

From: Portner, Ginny
Sent: Friday, October 04, 2002 11:10 AM
To: STIC-Biotech/ChemLib
Subject: 09/848909

Please search SEQ ID No 21 and homologs thereto. (protein claims) thanks,

Ginny Portner
CM1, Art Unit 1645
Room 7e13
Mail box 7e12
(703) 308-7543

Interference done as
well 10/02
vso

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

6329186
SEQ 8

5591631 SEQ 12
5677274 SEQ 12
5677274 SEQ 31

PCT/US 9401624
SEQ 12

acc #
AAR 60193
AAR 60183

WO 9418332

1994

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 10/4/02
Date Completed: 10/4/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 1
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: PR
WWW/Internet: _____
Other (specify): _____

PT or HIV-infected cells

PS Disclosure; Page 81-83; 12app; English.

CC The sequence encoding the protective antigen of *Bacillus anthracis*
CC may be used in the construction of a nucleic acid which encodes a
CC fusion protein comprising the anthrax protective antigen binding
CC domain of the native anthrax lethal factor and a sequence encoding an
CC activity inducing domain of a second protein. The fusion proteins
CC are useful for the specific killing of tumour cells or the killing of
CC cells infected with intracellular pathogens, especially HIV.

XX Sequence 735 AA:

Query Match 99.9%; Score 3774; DB 15; Length 735;
Best Local Similarity 100.0%; Pred. No. 4.2e-245;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 EVKQENRLNSESSESSOGILGYFSDLNFOAPWVWVSTTGDLSPSELENIPSENOYF 60
DB 1 evkqenrlnesesssgillgyfslndfnqapwvwtsttgdslspselempengyf 60
QY 61 OSATMSGFIKKKSDYEYTFATSDNHNWVWDOEYINKASNSKRTLEKGLYQIKIY 120
DB 61 osatmsgfiikkksdeytfatsadnhvwmvddgevinhasnskhrltkgylqiklqy 120
QY 121 GREMPTEKGLDFKLYTDSQNKKEVYSSDNLQPELKOKSSNRKRKRSISAGTYPPDRN 180
DB 121 grempetekglfdklytcdsqnkkevissdnlqpelkqssnsrkrkstisagtpvdrdn 180
QY 181 DGIPDSLEVEGYTVDVKNKRTLSFWISNHEKKGJLTJKSSPERKSTASDPSEKVT 240
DB 181 dgipdsleveytvdvknkrtlspwlsnhekkgjltksspekstasdpsefektv 240
QY 241 GRIDKNVSPRARHPLVAAPYIVHDMENILISKNEQSQNTDSEFTTISKVTSRTHT 300
DB 241 gridknvspearhplvaapyivhvmenilisknedqstqntdsefttisktsrht 300
QY 301 SEVHNAEHAASFDPIDGVSAGFSNNSSTVAIDHSLSIAGERTMAETGMLNTADTARL 360
DB 301 sevhnahaeasfdpiddgvsagfnsnstvaldhsislagertmaetgmlntadtarl 360
QY 361 NANIRRYNNGTAPYINVLPTTSLVYGNKQTLATIKAKENQSQILIANNNYPSKMLAPIA 420
DB 361 nanirrynngtapynvlpttslvynkqtlatikakengsqilapnnypskmlapi 420
QY 421 LNAQDPSSTPTNNYNOPELEKTKQRLDPOVYGNATYENRNRVNDGSMSEV 480
DB 421 lnaqdpsstptnnynopelkqrltdpovgnatyenrnrvndgsmsev 480
QY 481 LPOGFTKRIIFKNGKOLNVERRIAANVPSDPLETFKPDMLKEALKAFGEPEPNCNL 540
DB 481 lpgftkrlifngkdlmverriaavpsdpletfkpdmlealkafginepncnl 540
QY 541 OYQKDIETEDFNDOQTSQINKNOLAELNATNITYVLDKIKLAKAKNNILIDKRFPHDR 600
DB 541 oyqkdietedfnqdotsqinknolaelnatnityvldkiklakaknnilidkrfphdr 600
QY 601 NNTAVGADESYYKAEHREVINSSTEGILNIDNDIRKILISGTYIETDEGKLEKYNIRY 660
DB 601 nntavgaadesyykaehrevinsstegilnidndirkilistgyietdegkleyinry 660
QY 661 DMLNTSLRQDQKTFIDFKKYNKDLPLTISNPNYKYNVYAATKENTLINSNGDTSTNG 720
DB 661 dmlntslrqdqktfidfkynkdlpltisnppykynvyaatkentlinsngdstng 720
QY 721 IKRLIFSRRKGEIG 735
DB 721 ikrlifsrkgeig 735

```

RESULT 2

AAV56959
ID AAV56959 standard; protein; 736 AA.

AC AAV56959;

XX 25-APR-2000 (first entry)

DT B. anthracis MAT-PA protein.

DE B. anthracis MAT-PA protein.

XX *Bacillus anthracis* protein; protective antigen; PA; MAT-PA; TPA-PA;
XX tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS *Bacillus anthracis*.

PN WO200002522-A2.

XX 20-JAN-2000.

XX 09-JUL-1999; 99MO-US15568.

XX 10-JUL-1998; 98US-0092416.

PR (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

XX WPI; 2000-182165/16.

DR N-P-SDB; AA256875.

XX Recombinant DNA construct useful as vaccines for anthrax. In producing
XX host cells for analyzing the drugs and agents inhibiting anthrax

XX Disclosure; Page 34; 35pp; English.

CC The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of *Bacillus anthracis* proteins; selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis MAT-PA protein.

XX Sequence 736 AA:

Query Match 99.9%; Score 3774; DB 21; Length 736;
Best Local Similarity 100.0%; Pred. No. 4.2e-245;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 EVKQENRLNSESSESSOGILGYFSDLNFOAPWVWVSTTGDLSPSELENIPSENOYF 60
DB 1 evkqenrlnesesssgillgyfslndfnqapwvwtsttgdslspselempengyf 61
QY 61 OSATMSGFIKKKSDYEYTFATSDNHNWVWDOEYINKASNSKRTLEKGLYQIKIY 120
DB 61 osatmsgfiikkksdeytfatsadnhvwmvddgevinhasnskhrltkgylqiklqy 121
QY 121 GREMPTEKGLDFKLYTDSQNKKEVYSSDNLQPELKOKSSNRKRKRSISAGTYPPDRN 180
DB 121 grempetekglfdklytcdsqnkkevissdnlqpelkqssnsrkrkstisagtpvdrdn 181
QY 181 DGIPDSLEVEGYTVDVKNKRTLSFWISNHEKKGJLTJKSSPERKSTASDPSEKVT 240
DB 181 dgipdsleveytvdvknkrtlspwlsnhekkgjltksspekstasdpsefektv 241
QY 241 GRIDKNVSPRARHPLVAAPYIVHDMENILISKNEQSQNTDSEFTTISKVTSRTHT 300
DB 241 gridknvspearhplvaapyivhvmenilisknedqstqntdsefttisktsrht 301

```

```

OY 301 SEVHGNAEVHASFEDIGGSVSAAGFSNSNSTVAIDHSLSLAGEFTVAETMGINTADTARL 360
    |||||||
DB 302 sevbgnaevhasffidigsvsagfsnsstvaldhsislageretvaetmgintadtarl 361
OY 361 NANIRVYNTGAPLYNVLPFTSLVIGKNOTLATIRAKENOSQILAPNNVYPSKMLAPIA 420
    |||||||
DB 362 naniryvntgcaplynvlpftslvlgknotlatikakenqslqilapnnypskmlapia 421
OY 421 LNAODFESSPTITMNYNOFLEETKQRLDTPQVGNATVFNENGRVVDGTSNMSV 480
    |||||||
DB 422 lnaodfessptitmnynofleektkqrlidtpqvgnatvfnengrvvdtgnsnv 481
OY 481 LPDIOETTARIIFNGKDLNVERRIAANVSDLETPEDMTLEAKLAFENPNP 540
    |||||||
DB 482 lpdioettarilifngkdlnveritiaavnsdpeltpeamtlikealfinepn 541
OY 541 QYQGRDITEFDNFDOOTSNIKNOELNATVIVLDKIKLAKNNILIDKRFHYDR 600
    |||||||
DB 542 qyqgkditefdnfdqtsqnlknglaelnatlvivldkiklnaknnilidkrfhydr 601
OY 601 NNTAVGADSVYKAEHREVINSTEGLLNIDKIRKILSGYIEDEEGLEKEVINDRY 660
    |||||||
DB 602 nntavgadsvvkaehrevinsstegllnidkirkilsgylveledceglekevin 661
OY 661 DMLNISTLRQDGKTFIDFKKYNDKLPYISNPYKVNVAATKENTLINPSENGDTSTNG 720
    |||||||
DB 662 dmlnistlrqdgktdfkkyndkplyisnpnykvnyavatkentlinpsengdstng 721
OY 721 IKKILIFSKKGYELG 735
    |||||||
DB 722 ikkilifskkyelg 736

RESULT 3
AAY56960 standard; protein: 763 AA.
ID AAY56960
AC AAY56960;
XX
XX 25-APR-2000 (first entry)
DE B. anthracis TPA-PA protein.
XX
XX Bacillus anthracis protein; protective antigen: PA; MAT-PA; TPA-PA;
KV tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX
OS Bacillus anthracis.
XX
XX MO200002522-A2.
XX
XX 20-JAN-2000.
XX
XX 09-JUL-1999; 99WO-0515568.
XX
XX 10-JUL-1998; 98US-0092416.
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
PI WPI; 2000-182165/16.
XX
XX N-PSDB; AAY56876.
XX
XX Recombinant DNA construct useful as vaccines for anthrax. In producing
PT host cells for analyzing the drugs and agents inhibiting anthrax
XX
XX Disclosure; Page 32; 35pp; English.
XX
XX The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of Bacillus anthracis proteins, selected from protective
CC antigen (Pa), MAT-PA (PA with its secretory signals removed), TPA-PA

```

```

CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis TPA-PA protein.
XX
XX Sequence 763 AA:
SO
Query Match 99.9%; Score 3774; DB 21; Length 763;
Best Local Similarity 100.0%; Pred. No. 4,4e-245;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EYKQENRLNENSSSOGILCYESDLPAPAPVYTSSTDDLSISSRLNISEKNY 60
    |||||||
DB 29 eykqenrlnesessogilgyrfsdlnrqapmvysstgdsipsselenipsengyt 88
OY 61 QSAIVSGFTKKKSDERTFATSADNHYTMVDOEYINKASNSRKRLKGRLYQIKIY 120
    |||||||
DB 89 qsalvsgftkkksdertyatsadnhvmdvddqeylnkasnsrkrlkgrlyqikly 148
OY 121 QRENPTKGLDFKLYWTDSONKREYISSDNLOQLPELKOKSSNSKRRSTSAQPIVPRDN 180
    |||||||
DB 149 qrenptekgldfklywtdsqnkkevissdnlqlpelkksnsrkrrstsaqpi 208
OY 181 DGIPDLEEGYTVYVKNRRTPLSPWISNIHEKKGILTKYKSSPEKSTASDPSDEKVT 240
    |||||||
DB 209 dgipdsleegytyvdknrtrtlspwlsnhkkgilkyksspekstasdpdsflev 268
OY 241 GRIDKNVPEEARHPLVAAPYIVHDMENILISKEDOSTONTDEPTFKTSKTSRHT 300
    |||||||
DB 269 gridknvpearhplvaayplvhdmenniliskedgstntdsckrtsktsrht 328
OY 301 SEVHGNAEVHASFEDIGGSVSAAGFSNSNSTVAIDHSLSLAGEFTVAETMGINTADTARL 360
    |||||||
DB 329 sevbgnaevhasffidigsvsagfsnsstvaldhsislageretvaetmgintadtarl 388
OY 361 NANIRVYNTGAPLYNVLPFTSLVIGKNOTLATIRAKENOSQILAPNNVYPSKMLAPIA 420
    |||||||
DB 389 naniryvntgcaplynvlpftslvlgknotlatikakenqslqilapnnypskmlapia 448
OY 421 LNAODFESSPTITMNYNOFLEETKQRLDTPQVGNATVFNENGRVVDGTSNMSV 480
    |||||||
DB 449 lnaodfessptitmnynofleektkqrlidtpqvgnatvfnengrvvdtgnsnv 508
OY 481 LPDIOETTARIIFNGKDLNVERRIAANVSDLETPEDMTLEAKLAFENPNP 540
    |||||||
DB 509 lpdioettarilifngkdlnveritiaavnsdpeltpeamtlikealfinepn 568
OY 541 QYQGRDITEFDNFDOOTSNIKNOELNATVIVLDKIKLAKNNILIDKRFHYDR 600
    |||||||
DB 569 qyqgkditefdnfdqtsqnlknglaelnatlvivldkiklnaknnilidkrfhydr 628
OY 601 NNTAVGADSVYKAEHREVINSTEGLLNIDKIRKILSGYIEDEEGLEKEVINDRY 660
    |||||||
DB 629 nntavgadsvvkaehrevinsstegllnidkirkilsgylveledceglekevin 688
OY 661 DMLNISTLRQDGKTFIDFKKYNDKLPYISNPYKVNVAATKENTLINPSENGDTSTNG 720
    |||||||
DB 689 dmlnistlrqdgktdfkkyndkplyisnpnykvnyavatkentlinpsengdstng 748
OY 721 IKKILIFSKKGYELG 735
    |||||||
DB 749 ikkilifskkyelg 763

RESULT 4
AAY56958 standard; protein: 764 AA.
ID AAY56958
XX

```

AC	AAY56958;
XX	
DT	25-APR-2000 (first entry)
DE	B. anthracis protective antigen (PA) protein.
XX	
KM	Bacillus anthracis protein; protective antigen; PA; Mat-PA; TPA-PA; tissue plasminogen activator; PA65; vaccine; anthrax; antibacterial.
OS	Bacillus anthracis.
XX	
PN	MO200002522-A2.
PD	20-JAN-2000.
PF	09-JUL-1999; 99WO-US15568.
PR	10-JUL-1998; 98US-0092416.
XX	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PI	Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL. WPI: 2000-182165/16. N-Psdb; AAZ56874.
XX	
PJ	Recombinant DNA construct useful as vaccines for anthrax, in producing host cells for analyzing the drugs and agents inhibiting anthrax - disclosure; Page 33; 35pp; English.
PS	
CC	The invention provides a recombinant DNA construct that comprises a vector and at least one nucleic acid (or its fragment) encoding a combination of Bacillus anthracis proteins, selected from protective antigen (PA), Mat-PA (PA with its secretory signals removed), tPA-PA (PA with its secretory signals replaced with those of tissue plasminogen activator) and PA65 (63 kDa PA). The construct is useful as a DNA vaccine for anthrax and in producing infectious alpha virus particles. These particles, expressing the B. anthracis proteins are useful also as vaccines for anthrax. Host cells transformed with the construct are useful for analyzing the effectiveness of drugs and agents that inhibit anthrax or B. anthracis proteins. The present sequence represents a B. anthracis PA protein.
CC	
Sequence	764 AA:
Query Match	99.9%; Score 3174; DB 21; Length 764;
Best Local Similarity	100.0%; Pred. No. 4,4e-245;
Matches 735; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 EYKQENRLNLESESSSGGLAGYPSDLNQAPMVVTSSTTGDLSIPSELENIPSENQYF 60
DB	30 evkgenrlhnessessggllyyfsdlngapmvvtsttgldslpsseleinsenygf 89
OY	61 GSVAISGFIYKKSDIEYFAFSADNHVTMMVDDEVINKASNMKIRLENGRLYQIKIY 120
DB	90 gsalwsgfllvykkssdgyefatsadnhvtmvddevlnkasnmkirkylegrlyqikiy 149
OY	121 QRENPTEKGADFLWTDSQKKRVYSNNQLDELKOKSSNRKRKRTSAGTPYPDDN 180
DB	150 greopetkgjdfklwtidsqnkvkesonqlpelkqsasmrkrtksaagtpypddn 209
OY	181 DDPIPLSEVGTYGVYVKRKRTPLSPGISIHHEKGLTKYKSSPEKMNSAADPSPERYT 240
DB	210 ddplplslevgytygvvkxrflfspalshhekgyltkysppekmnsaadspdyseryt 269
OY	241 GRIDKNNSPARARPLVAAATPVYWDNETILISKEDOSTONTDSTRISKNTSTSRRHT 300
DB	270 gridknnspararplvaayplvhwmednliskkedgstontdstriksntstsrrht 329
OY	301 SEVHGNAEHAASFDDIGSVSAQSNSSSTVAIDHSLSLAGEKTMAETMGLTADPARL 360
DB	331 sevhnghaeahsafddigsvsaqsnssstvaldhslslagektmaetmgltadparl 360

Accession	Protein	Gene	Location	Notes
D8	330	sevjbnagvna	sfidfgsvsaqifmsnsstvaldhs	lsjagerlwaetnglntadrl 389
QY	361	NANIRYNNCTAPVAINV	PTTSVLVK	NOTLATIKAKENOLSOI:ASPNYYPSKNIAPLA 420
D8	390	nanirynnigtlapivn	pttsvlvknq	latikakenglsaq:lapnnyypskniapla 449
QY	421	LKAODSSSPITMNVN	OFLEETKTKOLALPD	OVGNATATYN:NSRPVNDGSWSEV 480
D8	450	lmgdftsscpimymn	gflelkkqlridtdq	vgylatcyntengvrvldgsnasev 509
QY	481	LPQIOETTAITRNK	OLNVERIAAVNSDP	LETTAKPMLK:ALKIAPGNEPNKL 540
D8	510	lpqietctatltnq	xdlvverriaavnsdp	letpkpmklxsklklafgnepnql 569
QY	541	QYQKDIETEDFNP	DOOTSINIKOLAE	LNATNITYVLCKILNAKNNILIRDRPHYR 600
D8	570	qyqskdletdfid	fgtqsgnlknqae	lnatnityvldklln:kmnlllrdkrfhydr 629
QY	601	NNIANGADESVY	KEAHREVINSSTBGL	LLNIDKDIRKLSGIV:ALDEPTGLEKVIYNDRY 660
D8	630	nniangaadesv	ykeahrevlnstseg	lllnidkdlrllfsgylvleldeglkxevlndry 689
QY	661	DMLNTSSLROD	KGTFIDKRYKNK	OLPLYSNPKNVAVYKEDU:LIINSENGOSTNG 720
D8	690	dmlntslrgdgtf	idfkkyknkplly	lspnpyknyvaykkel:llnpsengdstng 749
QY	721	IKKILIRFSK	KGYEIG 735	
D8	750	lkkilifskkyeig	764	
RESULT 5				
ABAB7306				
ID	ABAB7306	standard: Protein: 764 AA.		
AC	ABAB7306:			
DT	29-AUG-2001	(first entry)		
DE		wild type B. anthracis protective antigen.		
KW		lethal factor; LF: immunogen; LF4: protective antigen; PA: DNA vaccine		
KM		humoral; cell-mediated; immune memory response.		
XX		Bacillus anthracis.		
XX				
FR	Key	Location/Qualifiers		
FT	Peptide	1..29		
FT		/label= Signal peptide		
FT	Protein	/note= "Not given in the specification"		
FT	Peptide	30..764		
FT		/label= PA		
FT		204..764		
FT		/label= pcPA		
XX	MO200145639-A2.			
XX	28-JUN-2001.			
XX	21-DEC-2000: 2000MO-US34912.			
XX	22-DEC-1999: 99US-011459.			
XX		(OHS) UNIV OHIO STATE RES FOUND.		
XX		(GALL/) GALLOWAY D R.		
XX		(MATE/) MATECZON A J.		
XX	Galloway DR, Mateczun AJ:			
XX	WPI: 2001-408540/43.			
XX	N-PSDB: AAC86016.			
XX				
XX	Protecting animal against lethal infection with Bacillus anthracis, by			

PT administering wild type or mutated form of Bacillus anthracis lethal
PT factor protein or its fragment or a nucleic acid encoding the mutated
protein -

PS Claim 5; Fig 2; 33pp; English.

CC This sequence shows the B. anthracis protective antigen (PA). An
CC immunogenic fragment of PA, pcPA, can be used to produce an immune
CC response which protects an animal against lethal injection with
CC Bacillus anthracis. DNA encoding the B. anthracis PA can be used
CC in conjunction with DNA encoding the lethal factor (LF) in a
CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
CC or fragment alone or in combination with a DNA encoding the PA protein
CC or its fragment, both components (humoral and cell-mediated) of the
CC immune system are stimulated, which results in longer term immune
CC memory response. The combined use of a mutated LF and PA gene or their
CC fragments results in a higher level of immune response, as judged by
CC overall serum antibody titers for LF and PA antigens, than the use of
CC either LF or PA genes in separate immunizations.

XX Sequence 764 AA;

Query Match 99.9%; Score 3774; DB 22; Length 764;
Best Local Similarity 100.0%; Pred. No. 4,4e-245;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVKQENRLNSESSESSOGLGYFFSDLNFOAPWVTSTTGDLSIPSELENIPSENOYF 60
DB 30 evqgenrllnsesssgllgyffsdlnfqapmvvtsttgdlsipselelenseqyf 89
OY 61 QSAIWSGFIKVKKSDYEYFATSADNHYTMWVDDQEVINKANSNKRILEKGLYIKIOY 120
DB 90 qsaIwsGfIkVkkSdEYfAtSaDnHyTmWvDdQeVInKasNsKRIleKGLyIkIoY 149
OY 121 QRENPFTEKGLDFLYTDSQNKKEVYISDNILQPELKOKSSNKRKSTSAQPYDPDRDN 180
DB 150 qrenpfteKglDfLyTdsQnKkEYIsDnIlQPeLkOkSSnKRkStSaQpYdPdrDn 209
OY 181 DGIPDSLEVEGYTYDVNKRFTLSPMISNTHKKGITKYSKSPKSTASDPSDFEYVT 240
DB 210 dgIpdsLeVeGyTyDvNkrFtLsPmIsNtHkKgItKysKsPkStAsDpsDfEYvt 269
OY 241 GRIDKRVSPEARHPLVAAPIYVNDMENILSKNEQOSTQWDETRTTSKRTSRRHT 300
DB 270 grIdKrvSpEaRHpLVaApIyVNdMeNiLsKNeQoSQtQwDeTrTtSkRtSRRht 329
OY 301 SEVHGAEVHASFIDIGSVSGFSNSNSTVAIDHSLSLAGEYTAETMGINTADTARL 360
DB 330 seVhGaEvHAsFIdIGsVsGfSnsNsSTvAIdHsLSlAGeYtAeTmGInTAdTArL 389
OY 361 NANIRRYVNTGAPTYNVLPPTSLVGNKQTLATIKAKENQLSQTILAPNNYPSKNLAFIA 420
DB 390 naNIRrYvNtGaPtyNVlPpTslVgNkQtlAtIKaKeNqLSqTilApNnyPsKnlAfIA 449
OY 421 LNAQDDFSSTPTWNVNQPLELEKTKQLRLDTDOVGNIAFYENGRVAVDTGSNMGSEV 480
DB 450 lnaQddFsStPtWnvNqPlELeKtKqLRLdtDovGniAFyENgRvAVdTGsNmGSeV 509
OY 481 LPQIOGTFRITFGKDLNVERRIAANPSDPLETTKPTDLKEAKTAFENFPNNL 540
DB 510 lPqIOgtFRITfGkDlNvERrIAAnPsDpLEtTKpTDLkEAKtAFeNfPnnL 569
OY 541 OYOGKRIEPEPNDOOTSNTKKNOLALNATVIVVDKTKIAAKNITLIDKRFHIDR 600
DB 570 oYogKRIePePNdOoTSnTKKnOLaLnAtViVvDkTKIAAKnITlIdKRFhIDr 629
OY 601 NNIAVGADESYYKAREYINSSTGLLNDIDRKILISGIVIEIDTGBLKEYINDRY 660
DB 630 nnIAvGaDeSyyKaReYInSstGLlNdIdRkIlISgIVIEIdTGBlKeYInDrY 689
OY 661 DMLNSSLRQDGKFTIDFKKRYNDKLPLYISNPMYKRVNVAATKNTIINPSENGDTSTNG 720
DB 690 dMlNsSLRqDGKfTIdfKkRyNdKlPlYIsNpMYkRvNvAATkNTIInpSeNgDtStNg 749

DB 690 dmlnsslrdgkftidfkkyndklplyisnpykvvayavakem:linpsengdstng 749
OY 721 IKKILIRSKKGYETIG 735
DB 750 IKKILIRSKRYEIG 764

RESULT 6
AAR60183
ID AAR60183 standard; Protein; 903 AA.

AA60183;

04-APR-1995 (first entry)

PA(1-725)-----human CD4 fusion protein coding sequence.

Anthrax: Bacillus anthracis; fusion protein; lethal factor;

protective antigen; cell killing; targeting; pathogen;

intracellular; HIV; human immunodeficiency virus; toxin;

pseudomonas; exotoxin.

Bacillus anthracis.

Homo sapiens.

WO9418332-A.

18-AUG-1994.

14-FEB-1994; 94MO-US01624.

12-FEB-1993; 93US-0021601.

25-JUN-1993; 93US-0082849.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Atora N, Klimpel K, Leppla SH, Nichols PJ, Slough Y;

WPI: 1994-279753/34.

N-PSDB; AAQ70184.

Nucleic acid encoding anthrax toxin fusion protein - useful for

targeting toxin to specific cells, eg for killing tumour cells

or HIV-infected cells

Disclosure: Page 100-103; 124pp; English.

Sequence 903 AA;

Query Match 98.7%; Score 3729.5; DB 15; Length 903;
Best Local Similarity 99.2%; Pred. No. 5.5e-242;
Matches 728; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

OY 1 EVKQENRLNSESSESSOGLGYFFSDLNFOAPWVTSTTGDLSIPSELENIPSENOYF 60
DB 1 evqgenrllnsesssgllgyffsdlnfqapmvvtsttgdlsipselelenseqyf 60
OY 61 QSAIWSGFIKVKKSDYEYFATSADNHYTMWVDDQEVINKANSNKRILEKRLQIKIOY 120
DB 61 qsaIwsGfIkVkkSdEYfAtSaDnHyTmWvDdQeVInKasNsKRIleKRLyIkIoY 120
OY 121 QRENPFTEKGLDFLYTDSQNKKEVYISDNILQPELKOKSSNKRKSTSAQPYDPDRDN 180
DB 121 qrenpfteKglDfLyTdsQnKkEYIsDnIlQPeLkOkSSnKRkStSaQpYdPdrDn 180

```

OY 161 DCTDSELEVEGYTVVYVYKPKRTFLSPWISNHEKKGLTKSSPEKNSASDPYSDEKYT 240
DB 161 d9pdslevegylvkvknkrtflspwismhckgltksspeknsasdpysdelyt 240
OY 241 GRIDKNSEPARHPLVAAPYPIVHDMENIILSKNEDOSTQNTDSEFTRTISKNTSTSEFT 300
DB 241 gridknseparhplvaapyplivhdmehiilsknedgstqntdsetrtlskntstseft 300
OY 301 SEVGNNAEVAHFPGIGSVAGFSNSNSTVAIDHSLSLAGEERTMAETGLTADTATL 360
DB 301 sevgnnaevahfpgigsvagfsnnsnstvaialdhsllageretmactatatl 360
OY 361 NANRYNVTGTPRIYVLPFTTSVLGKNOTLATIKAKENOLSOILAPNNYPSKNLAPIA 420
DB 361 nanrynvgtapiyvlpfttsvlgknotlatikakenglsqilapnnypsknlapia 420
OY 421 LNAODDESSPTIMYNOPIELEKTKOLRLDPOYGNIAITNFENGVRVMDGNSMSEY 480
DB 421 lnaoddesptlmynopielekctkqlrltdpoygniatynfengrvvmdgsmsesv 480
OY 481 LPOIETARIFENKDLNVERRIAANVPSDPLETTPKPMIAKAKIAPGNEPNGL 540
DB 481 lpgietarilfngkdlntverriaavpsdplettpkpmialakialapgnepngl 540
OY 541 OYOGKDTTFEPDNFTQOTSONIKNOLAEINATNTIYVLDKIKLNKNNIILRKRPHDR 600
DB 541 oyogkdttfepdnftqotsoniknqlaelnatntiylvldkiklnaknniilrkrphdr 600
OY 601 NNIAVGADESVKAEHREVINNSTEGLINIDKIDIRKILSGYVEIEDETEGLEVINDRY 660
DB 601 nniavgadesvkeahrevinssteqliindkidlrlkllsgyveiedetegleevindry 660
OY 661 DMLNSSLRODKTEPIEFKRYNDKLELYISNPKNYAVVAKENTIIINSENGDSTWG 720
DB 661 dmlnsslrdqktfidkkyndkplyisnpnykvavvakentiiinpsengdstwg 720
OY 721 IKRLI--IFSKG 731
DB 721 ikkllkvylgkkg 734

```

RESULT 7

AAR60193 standard: Protein; 719 AA.

```

AC AAR60193:
DT 04-APR-1995 (first entry)
DE Modified protective antigen of Bacillus anthracis.
KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KM protective antigen; cell killing; targeting; pathogen;
KX intracellular; HIV; human immunodeficiency virus; toxin.
OS Bacillus anthracis.
XX
XX MO9418332-A.
XX PD 18-AUG-1994.
XX PF 14-FEB-1994; 94MO-USO1624.
XX PR 12-FEB-1993; 93US-0021601.
XX PR 25-JUN-1993; 93US-0062849.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Atora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y.
XX WPI: 1994-279753/34.
XX DR N-PSDB, AAO70189.
XX

```

PT Nucleic acid encoding anthrax toxin fusion protein - useful for
 PR targeting toxin to specific cells, eg for killing tumour cells
 PS or HIV-infected cells

Example 6; Page 114-115; 124bp; English.

The sequence encoding the protective antigen of *Bacillus anthracis* may be used in the construction of a nucleic acid which encodes a fusion protein comprising the anthrax protective antigen binding domain of the native anthrax lethal factor and a sequence encoding an activity inducing domain of a second protein. Such fusion proteins are useful for the specific killing of tumour cells or the killing of cells infected with intracellular pathogens, especially HIV, depending on the second component. The protective antigen and other toxins require proteolytic cleavage to acquire activity. Since some cells infected with an intracellular pathogen possess an active protease with quite a narrow substrate specificity e.g. HIV, the intracellular pathogen specific protease site (See AAR60184-89). The protease in cells that are infected with an intracellular pathogen cleaves the modified toxin which is then rendered active and kills the cell. This sequence is a modified Bacillus anthracis protective antigen which has the amino acids originally at positions 102-171 replaced with the HIV protease cleavable sequence described in AAR60186.

Sequence 719 AA:

Query Match 95.2%; Score 3597; DB 15; Length 719;
 Best Local Similarity 95.5%; Pred: No. 3-26-233;
 Matches 706; Conservative 4; Mismatches 5; Indels 24; Gaps 2;

```

OY 1 EVKQENRLNSESSESSQGLGYFEDLNFOAPMVYSSITGDLISSELENIPSENOYE 60
DB 1 evkqenrlnesessqgllyfyfedlnfqapmvvssitgdlisesselenipengyf 60
OY 61 QSAVNGCFKVKKSEDEYFATSADNHVYMWVDDQEVYKASNSKIRLEKGLYQIRKQY 120
DB 61 qsavngcfkvykksedeyfatsadnhvymvddqevylkasnskilrlekgllylqky 120
OY 121 QRENPEKGLDFKXKPTDSQNKKEYISDNLQPELKQKSSNS----RKRSTSGPTVP 176
DB 121 qrenpekgldfkxkptdsqnkkeyissdnlqpelkqkssns-----rkrstsgptvp 176
OY 121 qrenpekgldfklycdsqnkkeyissdnlqpelkqkssnsdalmmqagntfagpvp 180
DB 121 qrenpekgldfklycdsqnkkeyissdnlqpelkqkssnsdalmmqagntfagpvp 180
OY 177 DRDNDGIPDSLEVBGYTVVYVYKPKRTFLSPWISNHEKKGLTKYKSSPEKNSASDPYSDF 236
DB 177 drdndgipdslevegylvkvknkrtflspwismhckgltkyksspeknsasdpysdf 236
OY 181 drndgipdslevegylvkvknkrtflspwismhckgltkyksspeknsasdpysdf 240
DB 181 drndgipdslevegylvkvknkrtflspwismhckgltkyksspeknsasdpysdf 240
OY 237 EKVYGRIRKNVSEPARHPLVAAPYPIVHDMENIILSKNEDOSTQNTDSEFTRTISKNTS 296
DB 237 ekvygrirknvseparhplvaapyplivhdmehiilsknedostqntdsetrtlskntst 296
OY 241 ekvygridknvseparhplvaapyplivhdmehiilsknedgstqntdsetrtlskntst 300
DB 241 ekvygridknvseparhplvaapyplivhdmehiilsknedgstqntdsetrtlskntst 300
OY 297 RHTSEVGNNAEVAHFPGIGSVAGFSNSNSTVAIDHSLSLAGEERTMAETGLTADTATL 356
DB 297 rhtsevgnnaevahfpgigsvagfsnnsnstvaialdhsllageretmactatatl 356
OY 301 rhtsevgnaevahfpgigsvagfsnnsnstvaialdhsllageretmactatatl 360
DB 301 rhtsevgnaevahfpgigsvagfsnnsnstvaialdhsllageretmactatatl 360
OY 357 TARIANRIVYNGTAPIYVLPFTTSVLGKNOTLATIKAKENOLSOILAPNNYPSKNL 416
DB 357 tarianrivyngtapiyvlpfttsvlgknotlatikakenglsqilapnnypsknl 416
OY 361 tarianrivyngtapiyvlpfttsvlgknotlatikakenglsqilapnnypsknl 420
DB 361 tarianrivyngtapiyvlpfttsvlgknotlatikakenglsqilapnnypsknl 420
OY 417 APTALNAODDESSPTIMYNOPIELEKTKOLRLDPOYGNIAITNFENGVRVMDGNSMSEY 476
DB 417 aptalnaoddesptlmynopielekctkqlrltdpoygniatynfengrvvmdgsmsesv 476
OY 421 aptalnaoddesptlmynopielekctkqlrltdpoygniatynfengrvvmdgsmsesv 480
DB 421 aptalnaoddesptlmynopielekctkqlrltdpoygniatynfengrvvmdgsmsesv 480
OY 477 WSEVLPQIETARIFENKDLNVERRIAANVPSDPLETTPKPMIAKAKIAPGNEPNGL 536
DB 477 wsevlpqietaarilfngkdlntverriaavpsdplettpkpmialakialapgnepn 536
OY 481 wsevlpqietaarilfngkdlntverriaavpsdplettpkpmialakialapgnepn 540
DB 481 wsevlpqietaarilfngkdlntverriaavpsdplettpkpmialakialapgnepn 540
OY 537 NGMLQYOGKDTTFEPDNFTQOTSONIKNOLAEINATNTIYVLDKIKLNKNNIILRKRPHDR 596
DB 537 ngmlqyogkdttfepdnftqotsoniknqlaelnatntiylvldkiklnaknniilrkrph 596
OY 521 ngmlqyogkdttfepdnftqotsoniknqlaelnatntiylvldkiklnaknniilrkrph 580
DB 521 ngmlqyogkdttfepdnftqotsoniknqlaelnatntiylvldkiklnaknniilrkrph 580

```

OY 597 HYDNNIAGADESVYKHAHEVINSSTEGLLNIDKDIRKILSGYIVETEDTEGLKEVI 656
 DB 581 hydnniagadesvvykhahevinsstegllnldkdlrklisgyiveledteglkevi 640
 OY 657 NDYVDMNLNSSLRODGKTEFDKRYNDKLPYLISNKNVNVAVTEKNTIINPENGOT 716
 DB 641 ndyvdmlnsslrdgktdfkryndkplylispnkvnyvtevkentlinpengt 700
 OY 717 STNGIKRLIFSKKGYEIG 735
 DB 701 stngikrlilfskkyeig 719

RESULT 8

AAV56961
 ID AAV56961 standard; protein: 569 AA.

AC AAV56961;

DT 25-APR-2000 (first entry)

DE B. anthracis PA63 protein.

KM Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
 KW tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS Bacillus anthracis.

PN W0200002522-A2.

PD 20-JAN-2000.

PF 09-JUL-1999; 99WO-US15568.

PR 10-JUL-1998; 98US-0092416.

PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

DR WPI: 2000-182165/16.

DR N-PSDB; AAV56877.

PT Recombinant DNA construct useful as vaccines for anthrax, in producing
 host cells for analyzing the drugs and agents inhibiting anthrax

PS Disclosure: Page 35; 35pp; English.

CC The invention provides a recombinant DNA construct that comprises a
 vector and at least one nucleic acid (or its fragment) encoding a
 combination of Bacillus anthracis proteins, selected from protective
 antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 (PA with its secretory signals replaced with those of tissue plasminogen
 activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 for anthrax and in producing infectious alpha virus particles. These
 particles, expressing the B. anthracis proteins are useful also as
 vaccines for anthrax. Host cells transformed with the construct are
 useful for analyzing the effectiveness of drugs and agents that inhibit
 anthrax or B. anthracis proteins. The present sequence represents a
 B. anthracis PA63 protein.

SO Sequence: 569 AA;

Query Match 77.3%; Score 2919; DB 21; Length 569;
 Best Local Similarity 100.0%; Pred No 8, 2e-188;
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 STSAGPTVDDRNOCIDPDSLEVGCTVDYKRRKRTFLSPWISNHEKGLKRYSSPEKMS 227
 DB 2 stsagptvddrncidpdslevegtydvkrrkrtflspwlsnhelkrylksyspekms 61
 OY 228 TASDPYSDFEKKYGRKDNVSPARRHPLVAAYPIVHVDENITILSKNEDSTONTDSETR 287

DB 62 tasdpysdfekkygrdknvsparrhplvaaypihvdenitilsknedstqntdsctr 121
 OY 288 TISKNTSTSRHTSEVHNAEVIHASFEDIGSVSAGFSPNSNSVAADHSLSLAGEPTMA 347
 DB 122 tiskntstsrhtsevhnavehasffdigsvsagfnsnsvaidhsislagerewa 181
 OY 348 ETMGINTADYARLANIRRVNTGTAPTYNVLPTSLVGNKQIATKAKENOLSOILAP 407
 DB 182 etmgintadarlanirrvntgtaptynvlptslvgnkqiatkakenolsqilap 241
 OY 408 NNYPSKRLAPIALNODDPSPTNNYNQFLEKTKOLRLTPOOVGNATYNPENG 467
 DB 242 nnypskrlapialnoddpsptnnynqflektrkqlrltqygnlatyoteng 301
 OY 468 RRVNPTGSMNSEVLRIOETTRAKITFENGKDLNVEPRLAANP-38LEETRPDPTKEAL 527
 DB 302 rrvnptgsmnsevlrloettraktitfengkdlnveprlaanp-38leetrpdpptkeal 361
 OY 528 KIAFGNPNNGNLIQOGKITEFDPNFDOOTSONIKNOLAEINWITIVLDTKLNAK 587
 DB 362 kiafgnpgnngnliqgkltedfdpndootsoniknolaelnwityvldtklnak 421
 OY 588 NILINDKRFHYDRNNIAGADESVYKHAHEVINSSTEGLLNIDKDIRKILSGYIVE 647
 DB 422 nilindkrfhydrnniagadesvvykhahevinsstegllnldkdlrklisgyive 481
 OY 648 DTEGLKEVINDRYDMLNSSLRODGKTEFDKRYNDKLPYLISNKNVNVAVTEKNTI 707
 DB 482 dteglkevindrtdmlnsslrdgktdfkryndkplylispnkvnyvtevkentli 541
 OY 708 INPENGDTSTNGIKRLIFSKKGYEIG 735
 DB 542 inpengdstngikrlilfskkyeig 569

RESULT 9

AAE07903
 ID AAE07903 standard; Protein: 1052 AA.

AC AAE07903;

DT 01-NOV-2001 (first entry)

DE C. botulinum C2 translocation domain with Tent binding domain #2.

KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
 KW tumour; infection; neurodegenerative disease; gene therapy;
 KW botulinum neurotoxin; tetanus neurotoxin; Tent.

OS Clostridium botulinum.

OS Clostridium tetani.

PN W0200158936-A2.

PD 16-AUG-2001.

PR 04-DEC-2000; 2000WO-GB04644.

PR 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Shone CC, Sutton JM, Silman N;

DR WPI: 2001-514643/56.

PT New non toxic polypeptide for delivery of a therapeutic agent for the
 treatment of a CNS disorder comprising a binding domain that
 PT translocates the therapeutic agent into the neuronal cells -
 XX Example 2; Page 50; 50pp; English.

XX The invention relates to a non toxic polypeptide, for delivery of a
 CC therapeutic agent to a neuronal cell, which comprises a binding domain
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
 CC as HC) that binds to the neuronal cell and a translocation domain (amino
 CC terminal half of HC, designated as HN), that translocates the therapeutic
 CC agent into the neuronal cell, where the translocation domain is not a HN
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are
 CC useful for the treatment of a disease state associated with neuronal
 CC cells. The polypeptide constructs are useful for delivering therapeutic
 CC substances to neuronal cells. They are useful to treat disorders of the
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
 CC and infection. They are also useful in gene therapy. The present sequence
 CC is C. botulinum C2 enterotoxin translocation domain with tetanus
 CC neurotoxin (Tent) binding domain used in the exemplification of the
 CC invention.
 XX
 XX Sequence 1052 AA;

Query Match 22.6%; Score 855.5; DB 22; Length 1052;
 Best Local Similarity 33.1%; Pred. No. 9.5e-49;
 Matches 253; Conservative 138; Mismatches 297; Indels 77; Gaps 29;

QY 3 KOENRLNSES--SSQGLGTYPSDLNFOAPMWTSSSTGDSIPSSLENIPEENQYF 60
 DB 4 kfevsnksnknyftinglmgyfien-dfnlnlspcltdgnlftskedinslg-nkll 61
 QY 61 OSATMSGFIKXKSDTYFATSADN-HWTMWDQDEVINKASN-SNKIRLEKGLYQIKI 118
 DB 62 ksarwglkpsltyegylstnspcrvel--ngelflnstnsvnlhlgnydydlrl 118
 QY 119 -QYORENPTKGLD-FKLYWTDSONKKEVISSDNLQPELKOKSSNRKRSRSGPTV- 175
 DB 119 eqmnsenqllkyegiklywetadilkelpsevl-----lkpynsnekskfpimnlf 174
 QY 176 -----PRNDGCPDLEVEGTYVDVKNKRFELSPWTSNHKKGLPKYSSPEK 225
 DB 175 snaklkanaarldtdgldpewelingyvmnqkavawdkfcan-----gykkysnqpfk 229
 QY 226 WSTAPSPYSDPFKWTGRIDKNVSPKARHPLVAAPVPHVDMNITLSKNDGOSTONTSE 285
 DB 230 pctandpytdfkvsgqldpsvawaridpmisayplvgmerlvskse-----ltids 285
 QY 286 TRTISNTSTSTHSEVHGNAPVHASFQDGG-----SVSAGFNSNSSTVAIDSL 338
 DB 286 tkmsstshstnltv--gaevsgslqldagldfvyfmsasanyshcnqunslcvldl- 342
 QY 339 SLAGETAEVTEGANTADTARLNAIRVYNTGRAPVYNLPTTSGLVGNQTLATITAKE 398
 DB 343 --tge-stsqglstnhtgesaylmpilryngatapyntvptltvldk-qvavatkqge 398
 QY 399 NQLSQILAPNNYPSKMLADIALNMODFSSPTIYNNQFLEKTKQLRLDTDOVYGN 458
 DB 399 slldgylnggtyplldgppmalntndgfssrlplhnyqlksldngvmtstsqfign 458
 QY 459 IATVFNFGVAVDTGSNMSEVLPOIOTETARI--FNKDLNIVERRIAAVNPSPLET 516
 DB 459 fakyn-sngnlytvd-gnnwgyplgtlkstlslsfsgqtgya---vvaupfddedk 513
 QY 517 TRPDMLKAEALKIFGNEPENGMLQYOGKDT--EFDNEPDQOSTONIKNOQLAEINATN 573
 DB 514 t-pltlegalkafaleknkgyfjhgjleiskneklgyfldstnndfengqlktaakd 572
 QY 574 IYVTLKRIKLNKNNILIRDKRPHYDRNNIVAGDSVYKAEHREVNSTGELLNLDK 633
 DB 573 lmbcl--ikrn--mlllvkvtlkenisslnkldcwynededivllkksstlndldnn 628
 QY 634 DIRRIISGV---IYIEDTEGKLEVINRYDMNTNLSLKODKCFIDFKYNDKXLEPLIS 690
 DB 634 dilsdlsqfnsnvitlypdaglvpgingakahlnynesseyvthkamdi-eyndmf----- 692

QY 691 NPNKYNNYATKENTINPSENGDTSTNGIKILIFSCKGCEIG 735
 DB 683 -mftvstvlrvp---vsashlegylneysltsmkbhslsls 723

RESULT 10

AAE07901 standard; Protein; 1032 AA.

AAE07901;

01-NOV-2001 (first entry)

C. botulinum C2 translocation domain with Bont/F-binding domain #2.

Neuronal cell; binding domain; translocation domain; stroke; epilepsy;

tumour; infection; neurodegenerative disease; gene therapy;

botulinum neurotoxin type F; Bont/F.

Clostridium botulinum.

NO200158936-A2.

16-ANG-2001.

04-DEC-2000; 2000MO-GB04644.

02-DEC-1999; 99GB-0028530.

07-APR-2000; 2000GB-0008658.

(MICR) MICROBIOLOGICAL RES AUTHORITY.

Shone CC, Sutton JM, Silman N;

MP1: 2001-514643/56.

Example 2; Page 48; 50pp; English.

The invention relates to a non toxic polypeptide, for delivery of a
 CC therapeutic agent to a neuronal cell, which comprises a binding domain
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
 CC as HC) that binds to the neuronal cell and a translocation domain (amino
 CC terminal half of HC, designated as HN), that translocates the therapeutic
 CC agent into the neuronal cell, where the translocation domain is not a HN
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are
 CC useful for the treatment of a disease state associated with neuronal
 CC cells. The polypeptide constructs are useful for delivering therapeutic
 CC substances to neuronal cells. They are useful to treat disorders of the
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
 CC and infection. They are also useful in gene therapy. The present sequence
 CC is C. botulinum C2 enterotoxin translocation domain with botulinum
 CC neurotoxin type F (Bont/F) binding domain used in the exemplification of
 CC the invention.
 XX
 XX Sequence 1032 AA;

Query Match 22.6%; Score 854; DB 22; Length 1032;
 Best Local Similarity 32.8%; Pred. No. 1.2e-48;
 Matches 247; Conservative 139; Mismatches 259; Indels 108; Gaps 31;

QY 3 KOENRLNSES--SSQGLGTYPSDLNFOAPMWTSSSTGDSIPSSLENIPEENQYF 60
 DB 4 kfevsnksnknyftinglmgyfien-dfnlnlspcltdgnlftskedinslg-nkll 61
 QY 61 OSATMSGFIKXKSDTYFATSADN-HWTMWDQDEVINKASN-SNKIRLEKGLYQIKI 118
 DB 62 ksarwglkpsltyegylstnspcrvel--ngelflnstnsvnlhlgnydydlrl 118

```

OY 119 -OYOREMPTEKGLD-FKLYTDSOKKEVYSSDNLQPELKOKSSNRKRKSTAGPTV- 175
DB 119 eqlnseuqllknyegiklywetsdlkelpsevl---lkpysntnekskffpnltf 174
OY 176 -----PDRNDGIPDSLEVEGYVDVKNKRFELSPWISNIHEKKGLTKRKSSPEK 225
DB 175 snaklkanaanrtdtdgldpdeuwlnglytmqkavawdckfaan-----gykkvynpdk 229
OY 226 WSTASDPYSDPEKVTGRIDKNVSPPEARHPLVAAPYIVHVMENILISKNEQOSTONTDSE 285
DB 230 pclandpytdfakvsgldpsvsmvardpmsayplvgvmerlvyskse-----ltlqds 285
OY 286 TRTSKNTSTSRTHSEVHGNAEVAHASFEDICG-----SVASGFSNSNSTVAIDHSL 338
DB 286 tkmskstsstnlntv---gaevsgldagglfprfmsasanyshltwqntstvdct- 342
OY 339 SLAERTVAETGLMTADTARLANANRYVNGTAPITVNLPTTSVLAKNOTLATIKAKE 398
DB 343 --tge-sfsgslntgesaylmpnlrlyntgkavpynvptltlvdk-qvavtlkgge 398
OY 399 NQSLQILAPNNTYPSKNIAPALNADODSSPTITMNTNOFLELEKTKQLDLTDQYGN 458
DB 399 slldgylmpgtyplgeppmahcndqssrllpbynglksldngvtmstsgltgn 458
OY 459 IATYFENGVRVDTGNSMSEVLPQIOETTARIT--FNGKDLNVERRIAANPSDPLET 516
DB 459 fakyn-sngnlvtd--gnmwpylgtlksltasltstsgltqva---vvaaptsdpdk 513
OY 517 TKPDMTLEKALFIARGFNEPNGNIQYOGKDT--EPDNFDOOSTONIKQDLAELMNTN 573
DB 514 t-pltclgvalkafalekngkfyfhyjelskneklyqvlidshndfenglknadck 572
OY 574 IYTVLDIKILKNKNIILIRDRFRHYDRNNIAGVADSEVVEAHREYRSTEGLLNIDK 633
DB 573 lmbol--lkn--umllykvltf---kenis-----sintmgyndklllyfuk 615
OY 634 DIRRTLSGYVEIETEGELKEVINDRYDMANISSLRODCKTFIDFKKY-----NDLLET 687
DB 616 lykkikhsnld-----mryenankfidlsgysnltsgnyvyl 653
OY 688 YISNPNKVAVYATKENTITINSENGDSTNG 720
DB 654 ysturn-qfgly--skpsevnlaqndllyng 683

RESULT 11
AAE07900
ID AAE07900 standard: Protein: 1092 AA.
AC AAE07900:
DT 01-NOV-2001 (first entry)
XX
DE C. botulinum C2 translocation domain with BONT/F-binding domain #1.
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy;
KM botulinum neurotoxin type F; BONT/F.
XX
OS Clostridium botulinum.
XX
PN MO20015893-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000MO-G804644.
XX
PR 02-DEC-1999; 99GB-0028530.
XX
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MCCR-) MICROBIOLOGICAL RES. AUTHORITY.
XX

```

```

PI Shone CC, Sutton JM, Silman N;
XX
XX WPI: 2001-514643/56.
DR
XX
XX New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 2; Page 47; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as HC) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for treat disorders of the
CC CNS including neurodegenerative diseases; stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is C. botulinum C2 enterotoxin translocation domain with botulinum
CC neurotoxin type F (BONT/F) binding domain used in the exemplification of
CC the invention.
XX
XX Sequence 1092 AA:
SQ

Query Match 22.5%; Score 850; DB 22; Length 1092;
Best Local Similarity 32.0%; Pred. No. 2;3e-48;
Matches 249; Conservative 150; Mismatches 280; Indels 100; Gaps 33;

OY 3 KOEURLNSES--SSGILGYFSDNLFOAPMVTSTTGDLSPSELENIPISSNOVF 60
DB 4 kfensvsnknkyftinglmglyfen-dfnlnlspldgnltkskednslly-nkx1 61
OY 61 OSATWSGFIKVKKSDSEYTFATASN-HVTMWWDDQEVYKASN-SNKTRIEKRLQIKI 118
DB 62 ksatwylkpsltgelylstenpncrvel---ngelfnlstnscvmlngnyvdlrt 118
OY 119 -OYOREMPTEKGLD-FKLYTDSOKKEVYSSDNLQPELKOKSSNRKRKSTAGPTV- 175
DB 119 eqlnseuqllknyegiklywetsdlkelpsevl---lkpysntnekskffpnltf 174
OY 176 -----PDRNDGIPDSLEVEGYVDVKNKRFELSPWISNIHEKKGLTKRKSSPEK 225
DB 175 snaklkanaanrtdtdgldpdeuwlnglytmqkavawdckfaan-----gykkvynpdk 229
OY 226 WSTASDPYSDPEKVTGRIDKNVSPPEARHPLVAAPYIVHVMENILISKNEQOSTONTDSE 285
DB 230 pclandpytdfakvsgldpsvsmvardpmsayplvgvmerlvyskse-----ltlqds 285
OY 286 TRTSKNTSTSRTHSEVHGNAEVAHASFEDICG-----SVASGFSNSNSTVAIDHSL 338
DB 286 tkmskstsstnlntv---gaevsgldagglfprfmsasanyshltwqntstvdct- 342
OY 339 SLAERTVAETGLMTADTARLANANRYVNGTAPITVNLPTTSVLAKNOTLATIKAKE 398
DB 343 --tge-sfsgslntgesaylmpnlrlyntgkavpynvptltlvdk-qvavtlkgge 398
OY 399 NQSLQILAPNNTYPSKNIAPALNADODSSPTITMNTNOFLELEKTKQLDLTDQYGN 458
DB 399 slldgylmpgtyplgeppmahcndqssrllpbynglksldngvtmstsgltgn 458
OY 459 IATYFENGVRVDTGNSMSEVLPQIOETTARIT--FNGKDLNVERRIAANPSDPLET 516
DB 459 fakyn-sngnlvtd--gnmwpylgtlksltasltstsgltqva---vvaaptsdpdk 513
OY 517 TKPDMTLEKALFIARGFNEPNGNIQYOGKDT--EPDNFDOOSTONIKQDLAELMNTN 573
DB 514 t-pltclgvalkafalekngkfyfhyjelskneklyqvlidshndfenglknadck 572

```

OY 574 IYVLDKIKLNKNNILIRDKRHHYDRNNIAGVD-----ESVKEAHR-----YINSS 623
 DB 573 Imhcl-1ktn--mlllvkvtlkrnslnldnlfgyvsmrskvsgdgyraa 628
 OY 624 TEGLLNIDKDIR-----KILSGYIEIEDTGLKEVINDRYDL-----NISS 667
 DB 629 ttafsfk-skelkypgyymrftvqsyepflmsyundklllyfnklykklkdnslld 687
 OY 668 LRDKGKTFIDFKRY-----NDKLPLYSNPNRYKVNVAVENTENTINPSENGDSTNG 720
 DB 688 mysnakfkldisgysnslngvlysturn-qfgyly--sakpsvnaqnddllyng 743
 RESULT 12
 ID AAE07902 standard; Protein; 1112 AA.
 AC AAE07902;
 XX AAE07902;
 DE 01-NOV-2001 (first entry)
 DE C. botulinum C2 translocation domain with TeNT binding domain #1.
 KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
 KM tumor; infection; neurodegenerative disease; gene therapy;
 KM botulinum neurotoxin; tetanus neurotoxin; TeNT.
 OS Clostridium botulinum.
 OS Clostridium tetani.
 PN W0200158936-A2.
 XX 16-AUG-2001.
 PF 04-DEC-2000; 2000MO-CB04644.
 PR 02-DEC-1999; 99GB-0028530.
 PR 07-APR-2000; 2000GB-0008658.
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PI Shone CC, Sutton JM, Silman N;
 DR WPI; 2001-514643/56.
 XX New non toxic polypeptide for delivery of a therapeutic agent for the
 PT treatment of a CNS disorder comprising a binding domain that
 PT translocates the therapeutic agent into the neuronal cells -
 PS Example 2; Page 49; 50pp; English.
 CC The invention relates to a non toxic polypeptide, for delivery of a
 CC therapeutic agent to a neuronal cell, which comprises a binding domain
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
 CC as Hc) that binds to the neuronal cell and a translocation domain (amino
 CC terminal half of HC, designated as HN), that translocates the therapeutic
 CC agent into the neuronal cell, where the translocation domain is not a HN
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are
 CC useful for the treatment of a disease state associated with neuronal
 CC cells. The polypeptide constructs are useful for delivering therapeutic
 CC substances to neuronal cells. They are useful to treat disorders of the
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
 CC and infection. They are also useful in gene therapy. The present sequence
 CC is C. botulinum C2 enterotoxin translocation domain with tetanus
 CC neurotoxin (TeNT) binding domain used in the exemplification of the
 CC invention.
 XX Sequence 1112 AA;
 SQ

Best Local Similarity 32.18; Pred. No. 9e-48;
 Matches 240; Conservative 143; Mismatches 269; Indels 95; Gaps 30;
 OY 3 KOENRLNSES--SSQGLGYFSDNLFQAPMYVTSSTTGDS--ESSELENTPSENQYF 60
 DB 4 kfevsyvnshnkylftinglyfien-dflnlnlspfldgnlfskednslng-nkll 61
 OY 61 QSAIWSGFIVKVSDEYTFATSDN-HYTMVWDQEVINKASN-SNKIRLEKRLQIKI 118
 DB 62 ksarwglkpkltgelylstenpncrvel--pgelfnlslnstnylnlqgnvydirl 118
 OY 119 -QYRENPTEKGD-FKLYTWDSONKKEVISDMLOPELKOK-SNSKRRSTASAPRY 175
 DB 119 eqlnseuglkoyegiklywetsdlkelpsevl--lkpn/sntnekstipnltf 174
 OY 176 -----PRDNDGIPDSLEVEKGYVDVKNKRFELSPWISNHEKGLTKYKSSPK 225
 DB 175 snaklkannrdtdgldpdevelnglyymqkavaddkfaa-----gykkyvsnplk 229
 OY 226 WTSASDPTSDPEKVTGRIDKNVSPENRHPVAATPIYHDMENILSKNEDOSTOTDSE 285
 DB 220 pclandpytdfekvsgldpsvsmvwdpmsaypylygvmerlvwskse-----tlgds 285
 OY 286 TRTSKNPTSTSRHTSEVHGNAEVAHSPEDIG-----SVSAGFSNNSSTVAIDHSL 338
 DB 286 tkmskstshtslnlntv--gaevsqslqslgdlfpfmsaslnyhtwqstlvdd- 342
 OY 339 SLAGERTWAEIAGLNTADTARLNANIRYVNTGTAPINVLPTTSVZGRKQTLATIKAKE 398
 DB 343 -tge-sfsqslstlntgesaylnpnlryngltapyvnpitlvldk-qsvatlqge 398
 OY 399 NQLSQILAPNNYPSKKNLAPIALNADDFSSPTTMKNYNOFL-ETKQRLTLPDQVGN 458
 DB 399 slldgylnpggytlylgeppmalntndgfsrllipnqlksidngtvlatsqfqn 458
 OY 459 IATNFNGRVRVDGSMSEVLPOJQETARI--ENGKDLNVERIAAVPSPLET 516
 DB 459 fakyn-sngnlytd-gnnwgpplgltkettasltfsggttqva--vvaqpfsgpek 513
 OY 517 TKPMTKEKALKIAGFENPENGNOYOGKDIR--EFDNFPOQSONIKNOIAELMNTN 573
 DB 514 t-phltlegalkatalekngkfyfhglsknekqyflidsrlnndfengklndad 572
 OY 574 IYVLDKIKLNKNNILIRDKRHHYDRNNIAGVDESIVKEAHRVINSSTBGLLNDK 633
 DB 573 Imhcl-1ktn--mlllvkvtlkrnslnldnlfgyvsmrskvsgdgyraa 633
 OY 634 DIRKILSGYIEIEDTGLKEVINDRYDLNINSSLRDQK--FTI--DRKYNKPLV 688
 DB 609 ---qsmtqlsnrskqgdgyraatafsfkskelkypgyymrftvqsyepflmsyundk 665
 OY 668 ISNPNRYKVNVAVENTENTINPSENGD 715
 DB 666 vdn---eedldvllkstellindind 689
 RESULT 13
 ID AAM60224 standard; Protein; 880 AA.
 AC AAM60224;
 XX AAM60224;
 DE 28-SEP-1998 (first entry)
 DE Bacillus thuringiensis insecticidal toxin 177c8.
 KM Insecticide; pesticide; toxin; delta-endotoxin;
 KM biological control; lepidopteran; coleopteran.
 XX Bacillus thuringiensis strain PS17C8 (NRRL B-21667).
 OS
 FT Key Location/Qualifiers
 Misc-difference 253

```

FT      /note= "encoded by YTA"
FT      Misc-difference 675
FT      /note= "encoded by AC"
FT      Misc-difference 846
FT      /note= "encoded by RAA"
XX      MO9818932-A2.
XX      07-MAY-1998.
XX      30-OCT-1997; 97MO-US19804.
XX      30-OCT-1996; 96US-0029848.
XX      (MTCO ) MYCOGEN CORP.
XX      Dullum CJ, Feltelson JS, Loewer D, Muller-Cohn J,
XX      Narva KE, Schmeits JL, Schnepf HE, Schwab G, Stamp L,
XX      Stockhoff BA;
XX      WPI; 1998-272226/24.
XX      N-PSDB; AAV30307.
XX      Bacillus thuringiensis isolates - used for producing pesticidal
XX      toxins and nucleotide sequences for control of lepidopteran and
XX      coleopterans
XX      Claim 5; Page 81-84; 139pp; English.
XX      This polypeptide comprises a novel soluble toxin of Bacillus
XX      thuringiensis (B.t.) strain PS177C8 (NRRL B-21867). The toxin
XX      belongs to a novel family of B.t. toxins that have toxicity
XX      against non-mammalian pests. Its amino acid sequence was deduced
XX      from a novel DNA fragment (see AAV30307) obtained by PCR from
XX      cellular genomic DNA of PS177C8. Disclosed and claimed are novel
XX      B.t. isolates and toxins (see AAV60218-32) that have activity against
XX      lepidopteran and/or coleopteran pests. Isolated genes, probes
XX      and primers (see AAV30288-321 and AAV9734-87) useful for production
XX      of the toxins and for the identification and characterization of
XX      these toxins and transformed hosts, particularly plant and
XX      bacterial hosts. The invention provides 8 entirely new families of
XX      toxins from B.t. isolates. The toxins have the additional ability
XX      to form pores in cell membranes, and can be used to facilitate
XX      entry of a second agent into a target cell.
XX      Sequence 880 AA:
XX
Query Match 21.0%; Score 792.5; DB 19; Length 880;
Best Local Similarity 30.5%; Pred. No. 1.2e-44;
Matches 246; Conservative 134; Mismatches 298; Indels 129; Gaps 34;
OY      4 OENRLNESESSQGLGYFSLNFOAPVWVSTTGDLSPSELEN--IPSENOYFO 61
DB      39 qkng--qkmdrkgllgyfykygkdt-snlmfrprdstllygqfanklldkkgqyq 94
OY      62 SAISGFIKVRKSDDEYFATSNHVTWVDQEVINKASNSKIRLEKGRLOKIOYO 121
DB      95 slrvlgllysketgdtfnlsedeqatlelmgklnskkyvnhlekylpklkyq 154
OY      122 KEMPTKGLD-----FKLWTDSONKKEVYSSDNQLPELKQSS-----N 162
DB      155 sd--tkfnldsttkelklrkldasgdpqyqgqfndpefnkkesqeflakpsklnlft 212
OY      163 SRKRNSTAGPTVPDRNDGIPDSLEVBGTVDVKNKRRTFLSPMISNHEKKGLTKRYSS 222
DB      213 qkkrkrldeed--cdtdgdslpdlweengylf--qnlavkddsl-asgykykkyfsn 264
OY      223 PEKMTASDPYSDEKVTGRIDKNVSPARHPVAAVPIVHVDMENILISKNDSTONT 282
DB      265 pleshvvgdytydkaardldlslnaketfnplvaafpsvnmekyllspenlsls--- 320
OY      283 DSETRISKNTSTRTSTSEVGNALFVHASPDIIGSVAGPSNSNS--FVAIDHSLS 339

```

```

DB      321 ---nveshstmsyt-----ntc-----gasveaglygkysifgsvnyqshet 363
OY      340 LAGERTVAETMG-----LNTADTARLANIRIVNTGAPYVNLVTSVLVGRKNOTLATIK 395
DB      364 vage--wgtstgntsgfncasagylmanvrynvgfaydvkplcsfvl-ndclatlft 420
OY      396 AKENOLSOILAPNNYPSKKNLAPILANODPFSSTPTMNNYNQLEFEKTKOQLRDTDOV 455
DB      421 akenstalnispgesypkkygqgialtsmdfshhcltnkkyvdollnmkpmletnqt 480
OY      456 YGNMNTNPNENGRRVVDGSSNNSVLPQIOETARIIFNGKDLNVERRIAAVNSDPLE 515
DB      481 dg---vklctbgnitvrgemvngiygikaktsilvdgge-ivaskryakdyknped 536
OY      516 TTKPDMTLKEALAKAF--GENEPNGNLOYGKXDTLTFDF--NFVQJTSQNKINQIAEL-- 569
DB      537 kt-psltkdkdikstypdakegllyknkplysswmylydenlakevtkgindctg 595
OY      570 ---MATNITYLDRKIKLNAKNILIRDKRFRHPRNNIYAGADESVYKEAHREYINSSTEG 626
DB      596 kfkdvshlydv---kltpkmvnlk-lsilydn--aesndnslykwtntlvsgnng 647
OY      627 -----LLNLND-----KDIRRIISGYVIEDTE-----GLKEV 655
DB      648 kkyssnmpdanllnldaqeklnkryslsymkseknktgcclldgelypttktyrv 707
OY      656 INDYDMINT--SSLRODGKTFIDFKKYNKLPYLSNPNKRVNVAVTKEMTIINPSN 713
DB      708 nkdyrrldlalahniknpslshikt-ndetlflfddislt-dvasikpen--ltdeei 763
OY      714 GDF-STNGIK--KILFSPKKGVEIGZ 736
DB      764 kqysrygkledgllldkkgylhyge 790
XX
RESULT 14
XX      ID AAV59277 standard; Protein: 881 AA.
XX      AAV59277;
XX      18-APR-2000 (first entry)
XX      MIS toxin from B. thuringiensis strain PS177C8.
XX      Bacillus thuringiensis; toxin; endotoxin; pesticidal; plant pest;
XX      lepidopteran; colepterans.
XX      Bacillus thuringiensis.
XX      Key Location/Qualifiers
XX      Misc-difference 846
XX      /note= "unknown"
XX      MO9957282-A2.
XX      11-NOV-1999.
XX      06-MAY-1999; 99MO-US09997.
XX      06-MAY-1998; 98US-0073898.
XX      (MTCO ) MYCOGEN CORP.
XX      Feltelson JS, Schnepf HE, Narva KE, Stockhoff BA, Schmeits J,
XX      Loewer D, Dullum CJ, Muller-Cohn J, Stamp L, Morrill G,
XX      Flinstad-Lee S;
XX      WPI; 2000-096811/08.
XX      N-PSDB; AAV58786.
XX      New polynucleotides encoding pesticidally active proteins, useful for

```

transforming plants for controlling pests -
disclosure; Page 68-71; 104pp; English.

The invention relates to novel *B. thuringiensis* isolates, and genes encoding pesticidal toxins which are toxic to non-mammalian pests. The genes are useful in the control of non-mammalian pests and especially plant pests (e.g. lepidopterans and/or coleopterans). The polynucleotides are useful for transforming plants for controlling plant pests; for designing primers and probes useful for the identification and characterization of genes which encode pesticidal toxins. The present sequence represents a *B. t.* toxin.

SQ Sequence 881 AA;

Query Match	21.0%	Score 792;	DB 21;	Length 881;
Best Local Similarity	30.4%;	Pred. No. 1.4e-44;		
Matches 246;	Conservative 135;	Mismatches 297;	Indels 130;	Gaps 34

[illegible]

```

      1 11 111 11 11 1:
Db 764 ikqysrygikledgilldkkgghyge 791

```

RESULT	15
AA091239	
ID	AA091239 standard; Protein; 884 AA

DT	14-AUG-1996	(first entry)
XY		

DE B. cereus VIPIA(a) insect-specific protein
xy

KW Pesticide; insecticide; biological control agent; Lepidoptera;
KW Coleoptera; transgenic plant; maize; insect resistance;
KW western corn rootworm; *Diabrotica virgifera virgifera*; VIP.

OS Bacillus cereus strain AB78 (NRRL B-21058)
yy

Key	Location/Qualifiers
Reptide	1..33
	/label= Sig_peptide

PN W09610083-A1.

PD 04-APR-1996.

PF 27-SEP-1995; 95WO-EP03826

05-JUN-1995; 95US-0463483

XX

XX

PI KOZIEL MG, MULLINSM, NYE GJ, WARREN GW;

WPI; 1996-200921/20.

XXXXXX

growth - used in the control of Lepidoptera and Coleoptera pests.

PS Claim 15; Page 121-124; 242pp; English.

CC Insect-specific protein *VlPIA(a)*(AA091339) of *Bactillus cereus* AB78
CC shows activity against *Diatraea* spp. pests such as the
CC western corn rootworm. It is encoded by the *VlPIA(a)* gene (AA091340)
CC isolated from a cosmid clone of AB78. *VlPIA(a)* can be expressed in
CC e.g. bacterial hosts to provide biological control agents having
CC increased activity or target range, or can be expressed in transgenic
CC plants, esp. maize, to improve insect resistance. It is preferably
CC expressed as a fusion protein (see also AA091245) with auxiliary protein
CC *VlPIA(a)* (AA091238).

Sequence 884 AA;

Query Match	20.9%;	Score 790;	DB 17;	Length 884;
Best Local Similarity	30.4%;	Pred. No. 1.9e-44;		
Matches 248;	Conservative 130;	Mismatches 292;	Indels 146;	Gaps 34;

```

QY      4  OENRLHNESSSSOGILGIVFFEDLAFQAPMVMVSTTGGDLSIPSELEN--IRSENQGYO   61
Db      42  qnqng---gkemdtkjllgyyfyfkdf-sultmfaprcdrtlllydqatanklilkkgqey  97
QY      62  SAIVSGCFKVKAKDEVTPTATSDNHNVMWMDDEVINKNSNNKRLREKRLQIQTQYO   120
Db      98  stivtqlglsgkeqtgdftfnlsedegaaleingllsnkckekqvahlekglvprikley  157
QY     122  RENTPEKGID-----FKIYMWDSSNNKEVISSDNQLPLEIKOKSS-----N    167

```

QY 122 RNPTEKGLD-----FKLYWTDQNKKEVISSDNLQLPELKQSS-----N 16:

Db 158 sd--tkfnidskfkellfkidsqnpqvgdelrnpnfunkkeagflakpskinlft 215
OY 163 SRKRSTSAQPIPPDRNDGIPDSLEBYVDYVKNKRTFLSPMISNIHEKKGLTYKSS 222
Db 216 qkmkreided--tdtdgdspdlweenytl----gnriavwddsl-askytklfvsn 267
OY 223 PEKSTASDPYSPDKVGRIDKNVSPKARHPVIAAAYPIVDMENIILSKNEOSTMONT 282
Db 268 pleshltgdydcydekaardidlsnakefnplvaafpsvnmekvllspnenls---- 323
OY 283 DSERTISKTSRTSTSEVHGNAAEVAHAFSPDIGSVSAGFSNSNS--TVAIDHSLS 339
Db 324 ----nasvesnsstnswyt-----nte-----gasveagiqpkqisfgysvnyqhsct 366
OY 340 IAGERTVAETMG-----LNTADTARLANATRYVNTGTAPRYNVLPPTSVLGKNOTLATIK 395
Db 367 vage--wgsitqntsqfntasagylaanayvnytgaiydvkptctsfvl-ndtlatit 423
OY 396 AKENQLSQIILAPNPNYPSKNLAPIALNAODFSSPYTMANYNOFLELEKTKOLRLPTDOV 455
Db 424 aksnstalnispgeaypkpgqglaltsmndfnshpiltlnkkqvdllnokpmmlctngt 483
OY 456 YGNATATNFENGVRVDGNSWSEVLPOIOTETARIIFENGKOLNVERIRIAVNPSPDLE 515
Db 484 dg---vykkldhgnlvrgewngvlgkakstasiltvdage-ryaekrtvaakdyenped 539
OY 516 TTKPDMTKEALKIAF--GFNEPENGLOYOGKDTFEDF--NFDQTSQNIKNQLAEL-- 569
Db 540 kt-pelilkaalkisydpelkelegillyknpiyessvmcyldentakevckqjndltg 598
OY 570 ---NATNITVLDKTKLNAKMNILINDKRFHYDRNNIAGADSVVKEAREHYVNSSTEG 626
Db 599 kfkdvshlydv---klcpkmvrlk-lsilydn---aesndnsigkwtlnlnlvsqgnng 650
OY 627 -----LLNTND-----KDIRKILSGYVEIEDTE-----GLKE 654
Db 651 lkqysnnpdanhltntdageklknrdyislymkseknitqceltdagelyptlctvsn 710
OY 655 VINDRYDMLN-----ISSLRDQKTFIDPKKYNDKLPYISNPNYKVNVYAVTKE 704
Db 711 vnkdnykrldilahnikenpisslh-----lktndeiltfwdidsl-tdvasklpe 760
OY 705 NTIINPSENGDT-STNGIK--KILFESKGYEIGZ 736
Db 761 n--ltdseikqlysrqyikiedgilldkkgjlyhyge 794

Search completed: October 4, 2002, 14:35:58
Job time: 103 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 4, 2002, 14:34:15 ; Search time 24.37 Seconds

(without alignments)
2902.000 Million cell updates/sec

Title: US-09-848-909-21
Sequence: 3778
1 EVKQENRLNLSSESSQGLI...STNGIKILIFSKKGYEIG 736

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3774	99.9	764	2 139934	protective antigen
2	864	73.4	875	2 140862	lotia toxin compo
3	235.5	62.2	192	2 139933	cryptic protein -
4	235.5	6.2	204	2 G59104	hypothetical prote
5	199	5.3	4688	2 F82885	hypothetical prote
6	184.5	4.9	1227	2 C97033	uncharacterized pr
7	184.5	4.9	2401	2 T28676	rhoxy protein -
8	183.5	4.8	2529	2 B64635	toxln-like outer m
9	183	4.8	4152	2 T31102	filamentous hemag
10	182.5	4.8	1125	2 E90598	membrane nucleas
11	181	4.8	2269	2 T28677	rhoxy protein -
12	179.5	4.8	1639	2 S05603	major meroprote su
13	178.5	4.7	1072	2 A86827	hypothetical prote
14	178	4.7	752	2 G90589	hypothetical prote
15	178	4.7	1302	1 J06009	surface-located me
16	178	4.7	1658	2 S55101	hypothetical prote
17	178	4.7	2178	2 S55805	alpha-toxin - Clos
18	178	4.7	4919	2 T11105	hypothetical prote
19	178	4.7	6713	2 B89921	hypothetical prote
20	176	4.7	1193	2 S68218	hemolysin (impore
21	176	4.7	1635	2 A10452	hypothetical prote
22	173.5	4.6	821	2 S67087	hypothetical prote
23	173.5	4.6	1802	2 S52611	hemolysin (impore
24	172	4.6	1939	2 T18372	repeat organellar
25	171.5	4.5	1365	2 T30822	lmp1 protein - Myc
26	171.5	4.5	2399	2 H71879	toxln-like outer m
27	171	4.5	1465	2 S31262	TYB protein - years
28	171	4.5	1803	2 S56894	TYB protein - years
29	170	4.5	1837	2 T41023	probable nuclear p

30	169.5	4.5	786	2 T18469	hypothetical prote
31	169.5	4.5	1230	2 S56850	SMC1 protein homol
32	169.5	4.5	2340	2 B71704	cell surface anty
33	169	4.5	1308	2 E71622	probable membrane
34	168	4.4	1631	1 SAZOK1	major meroprote su
35	168	4.4	5005	2 F82884	hypothetical prote
36	167.5	4.4	1051	2 T18351	lmp1 protein - Myc
37	167	4.4	1315	2 T28679	fibriogen-binding
38	166	4.4	3724	2 T18427	hypothetical prote
39	165.5	4.4	979	2 J00894	p115 protein -
40	165.5	4.4	1115	2 T41342	probable coiled-co
41	165	4.4	3216	2 C90538	hypothetical prote
42	164.5	4.4	1467	2 PC1253	TYB protein - years
43	164	4.3	769	2 F89870	serine proteinase
44	164	4.3	1516	2 E71619	RA22 endonuclease
45	163.5	4.3	624	2 PC6003	surface membrane p

ALIGNMENTS

```

RESULT 1
139934
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence-revision 19-Jul-1996 #text-change 01-Dec-2000
C:Accession: I39934; S69160; F59104
R:Welkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodka, M.; Leppa, S.H.; Schmidt
Gene 69, 287-300, 1988
A>Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus
A:Reference number: I39933; MUID:8912073
A:Accession: I39934
A>Status: preliminary; translated from GB/EMBL/DDBI
A:Molecule type: DNA
A:Residues: 1-764 <RES>
A:Cross-references: GB:M22589; NID:9143280; PTDN:AAA22637.1; PID:9143282
R:Freeman, I.C.; Gordon, V.M.; Leppa, S.H.; Klimpel, K.R.; Birch, N.P.; Loh, Y.
Arch. Biochem. Biophys. 316, 5-13, 1995
A>Title: In vitro processing of anthrax toxin protective antigen by recombinant p
A:Reference number: S69160; MUID:95142670
A:Accession: S69160
A:Molecule type: protein
A:Residues: 197-202 <FRI>
R:Okinkwa, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Kelm, P.;
J. Bacteriol. 181, 6509-6515, 1999
A>Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid
A:Reference number: A59091; MUID:99445483
A:Accession: F59104
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313, 'O', 315-764 <OKI>
A:Cross-references: GB:AF065404; NID:94894216; PTDN:AO32414.1; PID:94894326
A:Experimental source: strain Sterne
A>Note: similar to anthrax toxin moiety, protective antigen, pagA formerly pag, p
C:Genetics:
A:Gene: pXOI-110
A:Genome: plasmid
C:Function:
A:Description: three component exotoxin; protective antigen binds to receptors on
y active components edema factor or lethal factor; the complex is internalized by
C:Keywords: exotoxin
F:1-29/Domain: signal sequence #status predicted <Sig>
F:30-196/Domain: propeptide #status predicted <Pro>
F:197-202/Product: protective antigen #status experimental <Mat>

Query Match 99.9% Score 3774; DB 2; length 764;
Best local similarity 100.0% Pred. No. 1.6e-183; indels 0; Gaps 0;
Matches 735; Conservative 0; Mismatches 0;
DB 30 EVKQENRLNLSSESSQGLI...STNGIKILIFSKKGYEIG 736

```

0y 61 OSATWGFATFKWKSSEPTKATSDNNTMMYMMWDDQEVYINKASNSKRTILEGRLQYIKIoy 120
 Db 90 OSATWSEFTIKVKSSEPTKATSDNNTMMYMMWDDQEVYINKASNSKRTILEGRLQYIKIoy 149
 0y 121 QRENPTFKGLDFRLYMTDSQNKREVISSDNQLPELKQSSNSKRKSTSAQPTVPDDND 180
 Db 150 QRENPTFKGLDFRLYMTDSQNKREVISSDNQLPELKQSSNSKRKSTSAQPTVPDDND 209
 0y 181 DQIPDSLEGVGYVYDVKNKRRTFLSPWISNIHEKKGLTKYKSSPEKMSSTADSPDEEYV 240
 Db 210 DQIPDSLEGVGYVYDVKNKRRTFLSPWISNIHEKKGLTKYKSSPEKMSSTADSPDEEYV 269
 0y 241 GRIQKNSVSPARHPVLAAYPIVHVDMENTILSKKEDOSTQNTDSEFTIKSTSRHT 300
 Db 270 GRIQKNSVSPARHPVLAAYPIVHVDMENTILSKKEDOSTQNTDSEFTIKSTSRHT 329
 0y 301 SEVHGNMVEVASPEQIGSVSAGFSNSNSTVYADHSLSLAGERTMAETGLMTADTARL 360
 Db 330 SEVHGNMVEVASPEQIGSVSAGFSNSNSTVYADHSLSLAGERTMAETGLMTADTARL 389
 0y 361 NNAIRVNTGTATVYVLTPTTSLYLGKNOTLATTKAENQSLIAPNNYTPSKNLAPIA 420
 Db 390 NNAIRVNTGTATVYVLTPTTSLYLGKNOTLATTKAENQSLIAPNNYTPSKNLAPIA 449
 0y 421 LNAODFSSPTITMANTNOFLELEKTKQLRLDPOYIGNIATYVPNGSVYRVDGSSMSRY 480
 Db 450 LNAODFSSPTITMANTNOFLELEKTKQLRLDPOYIGNIATYVPNGSVYRVDGSSMSRY 509
 0y 481 LPQIDETARIIRNGKDLNLYERIRIAVNPDSPELETTKPDMLTEALKIAGFENPGL 540
 Db 510 LPQIDETARIIRNGKDLNLYERIRIAVNPDSPELETTKPDMLTEALKIAGFENPGL 569
 0y 541 QYQCKRIITEPFENPDQSTQNKQNLAEIANNITVYLDKIKLNAMNILLRDKRHYDR 600
 Db 570 QYQCKRIITEPFENPDQSTQNKQNLAEIANNITVYLDKIKLNAMNILLRDKRHYDR 629
 0y 601 NNIAVGADESVYKARREYINSSTEGILLINDKIRKISGYIVELEPDEGKIVINDRY 660
 Db 630 NNIAVGADESVYKARREYINSSTEGILLINDKIRKISGYIVELEPDEGKIVINDRY 689
 0y 661 DMLNSSLKODGKTFIDFKKYNKDKPLIYISNPYKANNYAVTKEMLTINPSENDSTNG 720
 Db 690 DMLNSSLKODGKTFIDFKKYNKDKPLIYISNPYKANNYAVTKEMLTINPSENDSTNG 749
 0y 721 IKKILFSKKGYEIG 735
 Db 750 IKKILFSKKGYEIG 764
 RESULT 2
 140862
 Iota toxin component Ib - Clostridium perfringens
 C:Species: Clostridium perfringens
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
 R:Accession: I40862: S42774
 R:Perelle, S.; Gilbert, M.; Bogue, P.; Popoff, M.R.
 In:Act. Immun. 61, 5147-5156, 1993
 A:Title: Characterization of Clostridium perfringens Iota-toxin genes and expression in
 A:Reference number: I40862
 A:Accession: I40862
 A:Status: preliminary: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-875 <RES>
 A:Cross-references: EMBL:X73562; NID:g929031; PIDN:CA51960.1; PID:g414655
 Query Match 23.4% Score 884; DB 2; Length 875;
 Best local similarity 31.3%; Pred. No. 3, 5e-37;
 Matches 254; Conservative 131; Mismatches 276; Indels 150; Gaps 31;
 1 EYKQENRLNLESSSSGCLGTYFSDNLQAPMNVYVSTTGQGLSLPSSSELENIPSE-NQY 59

Query Match 6.2%, Score 235.5; DB 2; Length 192;

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
 A:Accession number: A96900; MUID:21359325; PMID:21359325
 A:Accession: C97033
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1227 <KUR>
 A:Cross-references: GB:AE001437; PID:NAK79054.1; PID:g15023995; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetic:
 A:Gene: CAC1080

Query Match 4.9%; Score 186; DB 2; Length 1227;
 Best Local Similarity 19.0%; Pred. No. 0.12;
 Matches 167; Conservative 16; Mismatches 287; Indels 288; Gaps 44;

```

37 SSTGDISPSELEINPSENOYFOSAIY-SGFTKKKSDSEYTFATSDMHTVMYVDDOE 95
14 SVTTSVALLSKPAIKAVADSVKSSISNSNENKGF-----VOEKE 58
96 V-INKASNSKIRLEKRLVQIKIYQRENPTEK-----GLDFKTYMDSQNKKEVI 146
59 YVWKKRNDNS- KVSSENQVS---NKNSNPKVSSSSSISQINKVNLQVQNNKSVL 114
147 SSDNLOLPKLOKSSNSRRK-----RSTSGPTVDDNDGIPDSLEYGYTVYNNKT 201
115 AASNVDEVAKKINSQVOTSDYIAIGETVYKEDLDLIINNAIDANSAG--TDLSEVEI 172
202 F-----LSPWISNI-----HERKG-LTKYK 220
173 YIYSQTAELAQAFRNINDGVANVSDYLLGATFVANDANLDSVKNKYFHKRYATVYK 232
221 SPSKRSKTA-----SDPYSPFVKYGRIDKNVSPRARHPLVAA 258
233 DVATTSALKNNINNGGEGDGYTALFVSGVQPYLD-----VKNKNYVEKO----- 280
259 YPIVAVNMENILISKMEOS--TQNTDEPTISK--NTSTRTSEVHGNFVHASFFDI 316
281 -----NKGRLDITELSDASTIARINFALDN-----MDA 311
317 GGSVSGNSNSSTYVAIDHSL-----SLAGERPMATM-----GLMTADTARLANI 364
312 GAVTLEDYQALGANVPOLEHAVNSLAMDORNGVSAIDGINTITINTINNIGVGT 371
365 -RYVNTGTAPI-----YVNLPTLSL--VLGKQTLATITKAKENQSLILANNY 411
372 DVIYNSHAVDSNEGNIDYDLNANITTEKTRAKGDL--TIPEVAVNYVEKVTLLDPEYHAA 430
412 -----PSKNLAPIALNADDFSSPTIMYNOFLEKTRKOLRLDTQV----- 455
431 AGOTTLQDKVNDPNAOYODDVAT-----LSMDLKTROCKTILKALQKIDSILNSLN 484
456 -----YGNINVTYFENGVRVDTGSMWSEVLPOLQETTARI--TENGKDLNVERRIAAVNP 510
485 INSGINIDYS-----KIQTEAVADAKLEAVNDIKIRAKGRDLTIOETROSVK 538
511 SDPLETT-----KPDWTLKELKIAF--GFNEPNGNLQYOGKRTDERFENDQOSTNKK 563
539 IDVINSTSVKSGDSYDITIGIDGYEI--NIEFVERIKREGITL--TINIKVAV 593
564 -----NOLATLN--ATNITYVLN-----KIKLNAMNLLIKRRFRHD 599
594 IEPYQJSEYRIYTGCVGVYDKTGLGINNVNOMNITINELKNKKDVK IODIOTFWD 653
600 R--NNT-----AVGADSEY-----VKAHREYINSPEGLLMDKDRILKLSYIVE 645
654 NTINNDVINKIGAGDAVLSDYNGIGITDYQDLDLVNADLKIONKVDODDIIEVAK 713
646 IEDTEGLK-----EVINDREYMLNISSKODGKFFIDFKKYNDKPLX-----ISNPKK 695
714 ISSYEALMRINIGEAVIDDFKALGLTDI-----NDGLLVAITTDLONNKYT 760
696 -----VWY-AVTKENTINPSENGDSTNGI 721

```

DB 761 ADEVIARVOAIEIYRALMQIN--LGRATYADYNTLGI 796

RESULT 7

728676

Rhopty protein - Plasmodium yoellii (fragment)

C:Species: Plasmodium yoellii

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000

C:Accession: T28676; A45521

R:Simla, R.A.; Keen, J.K.; Ogum, S.A.; Holder, A.A.

Mol. Biochem. Parasitol. 76: 329-332, 1996

A:Title: Comparison of two members of a multigene family coding for high-molecular

A:Accession number: 220507; MUID:97077455

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2401 <SIN>

A:Cross-references: EMBL:U06927; NID:g1041784; PID:AB41263.1

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42: 241-246, 1990

A:Title: Identification of the gene for a plasmodium yoellii rhopty protein. Multip

A:Reference number: A45521; MUID:91101660

A:Accession: A45521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 2260-2401 <KEE>

A:Cross-references: GB:M34281

Query Match 4.9%; Score 184.5; DB 2; Length 2401;
 Best Local Similarity 21.6%; Pred. No. 0.39;
 Matches 181; Conservative 120; Mismatches 316; Indels 221; Gaps 43;

```

1 EYKQENRLNESSESSQCLGYPYSDLN-----FOAPWVYTSSTTGDLST 45
402 EYAKENQVQLVYKSNLEIKKHNDQINIDNIKEKAKQVYDOFEKHKMTIPPENKRYOK 461
46 PSELEINPSENOYFOSAI--WSGFTKY--KPSDEYFATSAIMNVTMAYVDDOE--I 97
462 PSELEIKMDE--FLSKYKNYNDPOKVKYKEVSEHKKFTLNNKITEVSEDEIKRYE 518
98 NNASNSKIRLEKRLVQIKIYQRENPTEKGLDPLWYDSQNKKEVSSNLOLPELK 157
519 NFNDSKSLNETK--STEEYQNTITLKND--YIKQCLNTNELLITNHNKQTTLK 573
158 OKSSNRK--KRSTSGPTVDDNDGIP-----SLVEGYTVYNNKRTPLSP 205
574 DKINONIKTIEKENSIDKLYTDKFNELTDKTELETKFTGLSINNESNNKELLTYFD 633
206 WISNIHEKGLTYKSSPEKSTASDPYDFEYVGRIDKNVSPRARHPLVAAVYTVYVD 265
634 LKANLGRKNEMLYKQFNEK--EKAVEDIKKNDINKIVSN-----IET 677
266 MENILSKNEOSTQNTDEPTISKNTSTRTTSEVHGNFVHASFFDIGSAGS 325
678 IYTSIYNED--TENE--IGKSTIELNTRVLE-----KKAANT 713
326 NSNSSTVAI--DHSLSLAGERTVAETMGINTADTARLANIRVNTGTAPIYVNLPTSLV 384
714 NMEIKKIKYDQDFGK-----EKNIKYPDEN--KIKNDIDTLQOK 754
385 LGRN--OTLATIKR--KENQLSOI-----LAPN--NYVP--SKNLAPIA----- 420
755 IDSIEITTEIKSENHIDEIKQIDKLVKVPKTMFNEDEKLEKINIEYERIDKK 814
421 LNAODFSSPTIMYNOFLEKTRKOLRLDTDOYGNIAATYFENGVRVDTGSMWSE 479
815 NIKYKIDKLMKIEKINDKSTLEKKNINLSYKSGISMLFLQOVLIEKKRAHTIKKME 874
480 V--LPQIOETTARITFNGKDLN-----VERLIAVNSD-----LEFT-----K 518
875 AYIDDDINIKKRSOI--EKEMNIMDKIMDKIKEMKALNISHDQYKIYHTSKNHEK 931

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 14:34:50 ; Search time 13.55 Seconds

(without alignments)
2103.143 Million cell updates/sec

Title: US-09-848-909-21
Sequence: 1 EVKQENRLNSESSESSQGL.....STNGIKRLIFSKGVEIGZ 736

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3771	99.8	764	1	PAG_BACAN
2	235.5	6.2	204	1	YPL1_BACAN
3	179.5	4.8	1630	1	MSPI_PLAFC
4	179.5	4.8	1639	1	MSPI_PLAFC
5	178	4.7	1658	1	YMG7_YEAS
6	171	4.5	1803	1	YK13_YEAS
7	169.5	4.5	1230	1	SMC3_YEAS
8	163.5	4.3	2334	1	MARP_BACSU
9	162.5	4.3	978	1	PL15_MCHR
10	162	4.3	1957	1	YD86_SCHPO
11	161	4.3	1790	1	USO1_YEAS
12	160	4.2	1276	1	BKD_CLOBO
13	160	4.2	1385	1	FAP1_SCHPO
14	159	4.2	1000	1	S155_YEAS
15	156.5	4.1	1167	1	CAGA_HELPJ
16	155.5	4.1	1269	1	RBP1_PLAFC
17	155	4.1	1251	1	RBP2_PLAFC
18	155	4.1	1928	1	MYB1_YEAS
19	154.5	4.1	1024	1	RIP3_MOUSE
20	154.5	4.1	1208	1	PCP1_SCHPO
21	154.5	4.1	1726	1	MSPI_PLAFC
22	153.5	4.1	678	1	YMG7_YEAS
23	153.5	4.1	1487	1	MSB3_YEAS
24	153	4.0	1420	1	SRB3_YEAS
25	152.5	4.0	1116	1	YK54_AQUAE
26	151.5	4.0	1014	1	HEX2_YEAS
27	151.5	4.0	1225	1	Y109_MYCE
28	150.5	4.0	803	1	SMF6_YEAS
29	150.5	4.0	2116	1	MYG2_DICDI
30	149.5	4.0	1271	1	Y138_MYCE
31	149.5	4.0	1356	1	S1R4_YEAS
32	148.5	3.9	1577	1	HLRA_PROTI
33	148.5	3.9	1233	1	TF16_YEAS

34	148.5	3.9	1460	1	M159_YEAS
35	148.5	3.9	1728	1	MSPI_PLAFC
36	148.5	3.9	1744	1	TANA_XENIA
37	148	3.9	1570	1	PXK1_DICDI
38	147.5	3.9	1091	1	CIC2_RAT
39	147.5	3.9	1093	1	SM14_YEAS
40	147.5	3.9	2376	1	YMG7_YEAS
41	147	3.9	918	1	YMG7_YEAS
42	147	3.9	1288	1	VACA_HELPJ
43	146.5	3.9	1162	1	BXEN_CLOBO
44	146.5	3.9	1628	1	NACH_CLOPE
45	146.5	3.9	1916	1	RIP1_YEAS

ALIGNMENTS

RESULT 1	ID	PAG_BACAN	STANDARD	PRF	764	AA
AC	P13423	Q9P5R7	Q9KH69			
DT	01-JAN-1980	(Rel. 13, Created)				
DT	15-OCT-2001	(Rel. 40, Last sequence update)				
DT	01-MAR-2002	(Rel. 41, Last annotation update)				
DE	Protective antigen precursor (PA) (PA-83) (PA83) (Anthrax toxin translocating protein) [contains: PA-20 (PA20); PA-93 (PA93)]					P40477 saccharomyc
GN	PAG OR PAG OR PXO1-110					P04934 plasmidum
OS	Bacillus anthracis					001550 xenopus lae
OC	Bacteria; Firmicutes; Bacillus/Clostridium group					P54290 rattus norv
OC	Bacillus/Staphylococcus group; Bacillus					P25302 saccharomyc
ON	NCBI_TaxID=1392					P40468 saccharomyc
ON	11					P34487 caenorhabdi
RX	SEQUENCE FROM N.A.					092KX5 helicobacte
RA	MEDLINE-09172073; PubMed-3148491					P06366 clostridum
RA	Welkos S.L., Love J.R., Eden-McCutchan F., Vodka M., Leppia S.H., Schmidt J.J.					P26831 clostridum
RT	"Sequence and analysis of the DNA encoding protective antigen of Bacillus anthracis."					
RL	Gene 69:287-300(1988).					
RL	12					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-28_33, BA1024, AND BA1035					
RC	MEDLINE-09214082; PubMed-10197986					
RA	Price L.B., Hugh-Jones M., Jackson P.J., Keim P.					
RT	"Genetic diversity in the protective antigen gene of Bacillus anthracis."					
RL	J. Bacteriol. 181:2358-2362(1999).					
RL	131					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-V770-NP1-R / ATCC 14185					
RA	MEDLINE-20359347; PubMed-1089854					
RA	Cohen S., Mendelson I., Altboum Z., Kohler D., Jihany E., Blot T., Lettner M., Inbar I., Rosenberg H., Gores Y., Barak R., Fisher M., Kironman C., Velan B., Shafirman A.					
RT	"Attenuated nontoxicogenic and nonencapsulated recombinant Bacillus anthracis spore vaccines protect against anthrax."					
RL	Infect. Immun. 68:4549-4558(2000).					
RL	4					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-Sterne					
RC	MEDLINE-9445483; PubMed-10515943					
RA	Keim P., Koeber T.M., Lamke G., Kumano S., Mahillon J., Manter D., Martinez V., Riche D., Svensson R., Jackson P.J.					
RT	"Sequence and organization of pXO1, the large Bacillus anthracis plasmid harboring the Anthrax toxin genes."					
RL	J. Bacteriol. 181:6509-6515(1999).					
RL	151					
RP	DOMAINS					
RC	MEDLINE-91332080; PubMed-1651134					
RA	Singh V., Kimpel K.R., Quinn C.P., Chaudhary V.K., Leppia S.H.					
RT	"The carboxyl-terminal end of protective antigen is required for					

RT receptor binding and anthrax toxin activity.";
 RL J. Biol. Chem. 266:15493-15497(1991).
 [6]
 RA CHARACTERIZATION.
 RC STRAIN-Sterne;
 RX PubMed-8051159;
 RA Milne J.C., Furlong D., Hanna P.C., Wall J.S., Collier R.J.;
 RT "Anthrax protective antigen forms oligomers during intoxication of
 mammalian cells.";
 RL J. Biol. Chem. 269:20607-20612(1994).
 [7]
 RA CHARACTERIZATION.
 RC PubMed-11207581;
 RA Beauregard K.E., Collier R.J., Swanson J.A.;
 RT "Proteolytic activation of receptor-bound anthrax protective antigen
 on macrophages promotes its internalization.";
 RL Cell. Microbiol. 2:251-258(2000).
 [8]
 RA TOXIN REGULATION.
 RC STRAIN-Nepridge;
 RX PubMed-8300513;
 RA Koshner F.M., Dai Z., Kaufman-Yarbray M.;
 RT "Regulation of the Bacillus anthracis protective antigen gene: CO2 and
 a trans-acting element activate transcription from one of two
 promoters.";
 RL J. Bacteriol. 176:586-595(1994).
 [9]
 RA MUTAGENESIS OF PHE-342; PHE-343 AND ASP-344.
 RC STRAIN-Sterne;
 RX PubMed-7961869;
 RA Singh Y., Klimpel R.R., Arcora N., Sharma M., Leppla S.H.;
 RT "The chymotrypsin-sensitive site, PFD315, in anthrax toxin protective
 antigen is required for translocation of lethal factor.";
 RL J. Biol. Chem. 269:29039-29046(1994).
 [10]
 RA MUTAGENESIS OF DOMAIN 4 LOOPS.
 RC STRAIN-Sterne;
 RX PubMed-10085028;
 RA Vanughese M., Texeira A.V., Liu S., Leppla S.H.;
 RT "Identification of a receptor-binding region within domain 4 of the
 protective antigen component of anthrax toxin.";
 RL Infect. Immun. 67:1860-1865(1999).
 [11]
 RA MUTAGENESIS OF TRP-375; MET-379 AND LEU-381.
 RC STRAIN-Sterne;
 RX PubMed-1178978;
 RA Batra S., Gupta P., Chaudhan V., Singh A., Bhattacharjee R.;
 RT "Trp 346 and Leu 352 residues in protective antigen are required for
 the expression of anthrax lethal toxin activity.";
 RL Biochem. Biophys. Res. Commun. 281:186-192(2001).
 [12]
 RA MUTAGENESIS OF PHE-581; PHE-583; ILE-591; LEU-595 AND ILE-603.
 RC STRAIN-Sterne;
 RX PubMed-11554763;
 RA Ahuja N., Kumar P., Bhattacharjee R.;
 RT "Hydrophobic residues Phe552, Phe554, Ile562, Leu566, and Ile574 are
 required for oligomerization of anthrax protective antigen.";
 RL Biochem. Biophys. Res. Commun. 287:542-549(2001).
 [13]
 RA MUTAGENESIS OF PRO-289.
 RC STRAIN-Sterne;
 RX PubMed-11356563;
 RA Khanna H., Chopra A.P., Arora N., Chaudhary A., Singh Y.;
 RT "Role of residues constituting the 2beta1 strand of domain II in the
 biological activity of anthrax protective antigen.";
 RL FEBS Microbiol. Lett. 199:27-31(2001).
 [14]
 RA MUTAGENESIS OF GLN-512; ASP-541; LEU-543 AND ARG-621.
 RC PubMed-11222612;
 RA Mogridge J., Meunier M., Collier R.J.;
 RT "Involvement of domain 3 in oligomerization by the protective antigen
 moiety of anthrax toxin.";
 RL J. Bacteriol. 183:2111-2116(2001).

RM [15]
 RP MUTAGENESIS OF LYS-426; ASP-454 AND PHE-456.
 RX PubMed-11113126;
 RA Sellman B.R., Nassi S., Collier R.J.;
 RT "Point mutations in anthrax protective antigen that block
 translocation.";
 RL J. Biol. Chem. 276:8371-8376(2001).
 [16]
 RA X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RX MEDLINE-97192099; PubMed-9039918;
 RA Petosa C., Collier R.J., Klimpel R.R., Leppla S.H., Lidington R.C.;
 RT "Crystal structure of the anthrax toxin protective antigen.";
 RL Nature 385:833-838(1997).
 [17]
 RA REVIEW.
 RP PubMed-11544370;
 RX Mock M., Fouet A.;
 RT "Anthrax.";
 RL Annu. Rev. Microbiol. 55:647-671(2001).
 -1- FUNCTION: One of the three proteins composing the anthrax toxin,
 the agent which infects many mammalian species and that may cause
 death. PA binds to a receptor (ATR) in sensitive eukaryotic
 cells, thereby facilitating the translocation of the enzymatic
 toxin components, edema factor and lethal factor, across the
 target cell membrane. PA associated with LF causes death when
 injected, PA associated with EF produces edema. PA induces
 immunity to infection with anthrax.
 -1- SUBUNIT: Anthrax toxins are composed of three distinct proteins, a
 protective antigen (PA), a lethal factor (LF) and an edema factor
 (EF). None of these is toxic by itself. PA+LF forms the lethal
 toxin (Ltx); PA+EF forms the edema toxin (EdTx). PA-63 forms
 heptamers and this oligomerization is required for LF or EF
 binding. Once activated, at low pH, the heptamer undergoes
 conformational changes and converts from prepore to pore inserted
 in the membrane, forming cation-selective channels.
 -1- SUBCELLULAR LOCATION: Secreted.
 -1- DOMAIN: The molecule is folded into four functional domains. Each
 domain is required for a particular step in the toxicity process.
 Domain 1 contains two calcium ions and the proteolytic activation
 site. Cleavage of the PA monomer releases the subdomain 1a, which
 is the N-terminal fragment of 20-kDa (PA20). The subdomain 1b is
 part of the remaining 63-kDa fragment (PA63). Domain 2 is a beta-
 barrel core containing a large flexible loop that has been
 implicated in membrane insertion and pore formation. There is a
 chymotrypsin cleavage site in this loop that is required for
 toxicity. Domain 3 has a hydrophobic patch thought to be involved
 in protein-protein interactions. Domain 4 appears to be a separate
 domain and shows limited contact with the other three domains; it
 would swing out of the way during membrane insertion. Domain 4
 contains the binding sites for LF and EF and it is required for
 binding to the receptor; the small loop is involved in receptor
 recognition.
 -1- PTM: Proteolytic activation by furin or a furin-like protease
 cleaves the protein in two parts, PA-20 and PA-63; the latter is
 the mature protein. The cleavage occurs at the cell surface and
 probably in the serum of infected animals as well; both native and
 cleaved PA are able to bind to the cell receptor. The release of
 PA20 from the remaining receptor-bound PA63 exposes the binding
 site for EF and LF, and promotes oligomerization and
 internalization of the protein.
 -1- MISCELLANEOUS: In Ref. 9 multiple mutagenesis experiments were
 performed that showed that the residues present in the small loop
 of domain 4, and not the ones in the large loop, are involved in
 receptor recognition.
 -1- SIMILARITY: BELONGS TO THE BACTERIAL BINARY TOXIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

Query Match 99.8% Score 3771; DB 1; Length 764;
 Best Local Similarity 99.9%; Pred. No. 1 6e-181;
 Matches 734; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 EYKQERLLNNSSSSOGGLGTYEFDNLFOAPVYVTSSTGDLSTPSSELEMYENQYF 60
    |||||||
DB 30 EYKQERLLNNSSSSOGGLGTYEFDNLFOAPVYVTSSTGDLSTPSSELEMYENQYF 89
    |||||||
OY 61 QSALNSGFIKVKSDSEYTFATSDNHNVTMWDQEVINKASNSKIRLEKGLVQIKY 120
    |||||||
DB 90 QSALNSGFIKVKSDSEYTFATSDNHNVTMWDQEVINKASNSKIRLEKGLVQIKY 149
    |||||||
OY 121 QRENPTEKGLDFKLYMTDSQNKKEYISSDMQLPELKQSSNSKRRKSTSAQYVPRDN 180
    |||||||
DB 150 QRENPTEKGLDFKLYMTDSQNKKEYISSDMQLPELKQSSNSKRRKSTSAQYVPRDN 209
    |||||||
OY 181 DGIPSLSEEGYVDVKNKRTFLSPWISNHEKKGITYKSSPEKSTASDYSPEKVT 240
    |||||||
DB 210 DGIPSLSEEGYVDVKNKRTFLSPWISNHEKKGITYKSSPEKSTASDYSPEKVT 269
    |||||||
OY 241 GRIDNVSPEARHPVLAAYPIYHDMNITLSKNEDOSTONTDSERTSKSTSPRT 300
    |||||||
DB 270 GRIDNVSPEARHPVLAAYPIYHDMNITLSKNEDOSTONTDSERTSKSTSPRT 329
    |||||||
OY 301 SEYHGNAAVHASEFDIGSGVSGNSNSTVAIDHSLSLAGEKTAETMGANTPRARL 360
    |||||||
DB 330 SEYHGNAAVHASEFDIGSGVSGNSNSTVAIDHSLSLAGEKTAETMGANTPRARL 389
    |||||||
OY 361 NANIRYVTGAPVYVLEPTTSVLGKNOFLATIKAKENLSQILAPNNYPSKMLAPIA 420
    |||||||
DB 390 NANIRYVTGAPVYVLEPTTSVLGKNOFLATIKAKENLSQILAPNNYPSKMLAPIA 449
    |||||||
OY 421 LNAODFFSPTITMNYNOFLELEKTKOLRLDTDOYVGNIAATYFNENGRVAVDGSNSSEV 480
    |||||||
DB 450 LNAODFFSPTITMNYNOFLELEKTKOLRLDTDOYVGNIAATYFNENGRVAVDGSNSSEV 509
    |||||||
OY 481 LPDIQETARIIFNGKDLNVERRIAANPSDPLETTRPDTELEAKLAFENPNPNTL 540
    |||||||
DB 510 LPDIQETARIIFNGKDLNVERRIAANPSDPLETTRPDTELEAKLAFENPNPNTL 569
    |||||||
OY 541 QYOGKDITEFDFNFDOOTSQNKIKNOLAELNATNITYVLDKIKLAKANNILIDRKFPHDR 600
    |||||||
DB 570 QYOGKDITEFDFNFDOOTSQNKIKNOLAELNATNITYVLDKIKLAKANNILIDRKFPHDR 629
    |||||||
OY 601 NNINAVGADESVEKAREVINSSTEGILLNIDKRIKILSGIYEIEDTEGKKEVINDRY 660
    |||||||
DB 630 NNINAVGADESVEKAREVINSSTEGILLNIDKRIKILSGIYEIEDTEGKKEVINDRY 689
    |||||||
OY 661 DMITNSSLROGKTEFIDFKKYNDKPLIYISNPYKANYAATKRENTIINSPNGDSTNG 720
    |||||||
DB 690 DMITNSSLROGKTEFIDFKKYNDKPLIYISNPYKANYAATKRENTIINSPNGDSTNG 749
    |||||||
OY 721 IKKILIFSKKGYEIG 735
    |||||||
DB 750 IKKILIFSKKGYEIG 764
    |||||||
  
```

RESULT 2
 YPBI_BACAN STANDARD; PRT; 204 AA.
 AC P13422; Q9X377;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein px01-111.
 GN PX01-111.
 OS Bacillus anthracis.
 OC Plasmid px01.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacilli/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1392;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=89172073; PubMed=3148491;
 RA Melios S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppia S.H.,
 RA Schmidt J.U.;
 RT Sequence and analysis of the DNA encoding protective antigen of
 RT Bacillus anthracis.
 RL Gene 69:287-300(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Steinere;
 RA MEDLINE=99445483; PubMed=10515943;
 RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.E., Hill R.K.,
 RA Keim P., Koehler T.M., Lamme G., Kumano S., Mahillon J., Menter D.,
 RA Martinez Y., Rieke D., Svensson R., Jackson P.J.;
 RT "Sequence and organization of px01, the large Bacillus anthracis
 RT plasmid harboring the Anthrax toxin genes".
 RL J. Bacteriol. 181:6509-6515(1999).
 CC -i- SUBCELLULAR LOCATION: Membrane-bound (potential).
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M22897; AAA22636.1;
 DR EMBL: AF065404; MAD32415.1;
 DR HSSP: P13423; JACC.
 KW Hypothetical protein; Plasmid; Transmembrane.
 FT TRANSMEM 162 182
 FT CONFLICT 93 93 T->A (IN REF. 1).
 FT CONFLICT 184 204 KSCNCLLYVEVSQIMNSVEY -> NHVYILSM (IN
 FT REF. 1).
 SO SEQUENCE 204 AA; 23029 MW; E1657823AE4273FD CR=64;

Query Match 6.2% Score 235.5; DB 1; Length 204;
 Best Local Similarity 34.6%; Pred. No. 1 3e-05;
 Matches 56; Conservative 36; Mismatches 47; Indels 23; Gaps 6;

```

OY 587 MNLIRDRFRHYRNINAVGADESVEKAREVINSSTEGILLNIDKRIKILSGIYEI 646
    |||||
DB 1 MNLVFRDP-YHDMNGNIYGVDDSYLKNKAKQILNMSSDGVSLNLEDEVQALSGYMIQI 59
    |||||
OY 647 EDTE-----GLKEVINDRYDMNLSSLRGDKTEFIDFKKYNDKPLIISNP 693
    |||||
DB 60 KRPSNHLNNSPYITILAGDSGVBELRVLS-----DGTGTFDPKKPDEMNRSLV-DPG 112
    |||||
OY 694 YVNYVYATKEN-TIINSENDSTNGCIKILIFSKKGYEI 734
    |||||
DB 113 DQVYVYATKEDRNATYRDENGNIA-HKLNNTLVLSGKIKEI 153
    |||||
  
```

RESULT 3
 MSP1_PLAFK STANDARD; PRT; 1630 AA.
 AC P04932;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (Isolate K1 / Thailand).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86136024; PubMed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 RA Stuenkelberg H., Bujard H.;

	RT	"Polymorphism of the precursor for the major surface antigens of rat plasmidum falciparum mezozoites: studies at the genetic level."
	RN	[EMBO J. 4:3823-3829(1985).]
	RN	REVISTIONS. SEQUENCE FROM N.A.
	RA	Pan M., Tolle R., Bujard H.;
	RL	Submitted (JUN-1995) to the EMBL/GenBank/DDBI databases.
	CC	- I - SUBCELLULAR LOCATION: Attached to the membrane by a GPI anchor (potential).
	CC	- I - PMV: MEZOZOITE SURFACE ANTIGEN COMPAIN THE SEQUENCE OF 83 kDa, 42 kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEZOZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
	CC	OR send an email to license@lsb-sib.ch).
	CC	-----
	DR	EMBL; X03371; CAA27070.1; -
	DR	PIR; A25120; SAZOKI.
	DR	InterPro; IPRO000561; EGF-like.
	DR	Pfam; PF00008; EGF_1.
	KM	Malaria; Mezozoite; Polyprotein; Repeat; Signal; Glycoprotein;
	KM	Transmembrane; GPI-anchor.
	FT	SIGNAL 1 19 POTENTIAL.
	FT	CHAIN 20 1630 MEZOZOITE SURFACE PROTEIN 1.
	FT	DOMAIN 67 84 THIEPPTIDE SG(TP) REPEAT.
	FT	TRANSMEM 1614 1630 MEMBRANE ANCHOR.
	FT	CARBOHND 97 97 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 259 259 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 755 755 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 759 759 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 774 774 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 835 835 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 911 911 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 935 935 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 1049 1049 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 1156 1156 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 1165 1165 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 1436 1436 N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHND 1517 1517 N-LINKED (GLCNAC . .) (POTENTIAL).
	SQ	SEQUENCE 1630 AA; 187289 MW; ADDBEC3CE0M46322 CRC64;

		Query Match 4.8%; Score 179.5; DB 1; Length 1630;
		Best Local Similarity 20.4%; Pred.No.0.13;
		Matches 171; Conservative 111; Mismatches 272; Indels 285; Gaps 41.
Oy	5	ENRLTLNESSSSOGLILGYPSDLNFQAPMVYTSSTGDISIPSELENIPESENQFOSA 63
		::: ::: ::: ::: ::: ::: ::: :::
Ddb	918	EMLSLCKNNINYOELIGOKSESE-NF-----YEKLIKSDPFNYE 956
Oy	64	INSGIKKWKSDEYFPATSAHNVTMMVDDEVINKRASNS-----NKTRLEKGRLY- 114
		:: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Ddb	957	SPTFNVSXKAD-----INSLNDSKRKKILEEDIKTKTQLDLPDLRYNRYKTKLERFLD 1011
Oy	115	--OIKIOVOYREMPREKLDGPKLYWDSONKNKEYISDNULDELLOKSSNSKKRS 168
		: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Ddb	1012	KKKTYGVKKQIKKILLKEGLEKLT--NSLNMKHVL-QNFVSVPFRKKKELEIAETEN 1067
Oy	169	TSGAGTYPDNDNGIPDSLVEPGYTDVQNRKRTLSLWMISNHKKGLGTG---KSP-- 223
		::: ::: ::: ::: ::: ::: ::: :::
Ddb	1068	T-----LENRKILIKHY-----KGLIVTYMGESPLPK 1094
Oy	224	---EKNSTASDPISDFE-----KYTGRIIDKWSPEER-----HPLVAAIPIVHVD 265
		::: ::: ::: ::: ::: ::: ::: :::
Ddb	1095	TIISESIIOTBEDYNASLENFKYSLSGKLKDLMLEKKALSTISGGHHHLA----E 1147
Oy	266	MENTILSKNDOSTONDSETRTISKNTSRTPRSEVGNAEVHASFDIGGSVAGFS 325
		::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::

D	b	1148	LKEVTKRNN---YTSNSENNT-----DYNNALESKREPLEGEGDAVAYS	1191
Oy		326	NNSNSYADHSILSLAGEPMATMGLMTADTARLANNRVNGTPITNYLPTP----	381
D	b	1192	EESGDSTLESGQRKKRPASTHVGAS---NMITSQ--HYDEVEDDYIVPIGESEEDYDL	1247
Oy		382	-SVLGSNQIATIKAKENOLSOILAIPNNYPSSNLAPIALMAADQDSPPITMYNOFL	440
D	b	1248	GQVTGGAYTPSYT---DNLSKI---ENEDEVYLAKPELA-----GYIR	1285
Oy		441	ELEKTROLADDDOYGNTATFENGWY---RVDGSWSMSYLPJQGETARIIFNKD	497
D	b	1286	SLEK-----QLNNWFTEFNWVKDLSRNKRRENFKNVL-----SD	1323
Oy		498	L----NLVERRLAANVPSPDELTTTFKDTLKALKALAFGFNEPN--NLQOGKDITEPDFN	553
D	b	1324	LIPRKDLTSSNVVVDPKPLMKRKHDFLSYNYI-----DSDSIDIDIN	1368
Oy		554	PDOOT-----SONIKMOIAEL-----NAFNLY-TVLDKI-----SBL	
D	b	1369	FANDVLGYKYRISEKYEKSDISIKTYINDKGENEKRYEFPFLNIETLYKYNDKDIIEVI	1428
Oy		582	-----KLAKANNI-----LIRKRFPHDRNNIAVTADESVEKRAHEVINS	622
D	b	1429	HLEAKVANTAYESESVEYKIKELNLYLKTIOCKLADRKKNNVFVLIADLSTDYNNHNHLTK	1488
Oy		623	--STEGLILNDKDI-RKILSGYT--VEIEDFGALEVINRYNRMNISLRDGKTFPI	676
D	b	1489	PLTGHWFEMLAKVTLSMLDGNLOGMINTISOHCYKCCPONSSCPRHIDEPEGCKLT.	1548
Oy		677	DEKKYNDKLELVISNPXYXVAVTKENTIIENSE-----DTSGNIGIKTI	724
D	b	1549	NFKDGRKC---VENPNPTCN-----NGGCQADAKCTREDSSSGN-KKI	1590
RESULT 4				
MSP1_PLAIN				
ID MSP1_PLAIN	STANDARD:	PRF:	1639 AA.	
AC	FO04933:			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-FEB-1996 (Rel. 33, Last annotation update)			
DE	Mezozoit surface protein 1 precursor (Mezozoit surface antigens) (PUMSA) (P195).			
GN	MSP-1.			
OS	Plasmodium falciparum (isolate Wellcome).			
OC	Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5848;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=66014355; PubMed=2995820;			
RA	Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,			
RA	Nicholls S.C., Hillman Y., Davey L.S., Tizard M., V., Schwarz R.T.,			
RA	Freeman R.R.;			
R7	Primary structure of the precursor to the three major surface			
R7	antigens of Plasmodium falciparum mezozoites.*;			
RL	Nature 317:270-273(1985).			
RN	[2]			
RP	REVIEWS:			
RA	Holder A.A.;			
CC	Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.			
CC	!- SUPPLEMENTARY LOCATION: Attached to the membrane by a GPI-anchor			
CC	(Potential)			
CC	!- PRT: MEZOZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42			
CC	kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF			
CC	MEZOZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outpost at			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/)			

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; X02919; CAA2676.1; -.
DR PIR; A24594; A24594.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merzoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1639
FT CARBOHYD 116 116
FT CARBOHYD 268 268
FT CARBOHYD 764 764
FT CARBOHYD 768 768
FT CARBOHYD 783 783
FT CARBOHYD 844 844
FT CARBOHYD 920 920
FT CARBOHYD 964 964
FT CARBOHYD 1058 1058
FT CARBOHYD 1165 1165
FT CARBOHYD 1174 1174
FT CARBOHYD 1445 1445
FT CARBOHYD 1526 1526
SQ SEQUENCE
1639 AA; 187618 MW; 2C235B6616C87F6E CRC64;

Query Match 4.8%; Score 179.5; DB 1; Length 1639;
Best Local Similarity 20.4%; Pred. No. 0.13;
Matches 171; Conservative 111; Mismatches 272; Indels 285; Gaps 41;

OY 5 ENRL-LSNESSSOGGLGYFSDLNFOAPWVYTSYTGDLSPSSLELNPSENGYFOSA 63
DB 927 ENLISGKNKRYIOELIGKSSSE-NF-----YKILNDSTFYNE 965
OY 64 IMSGFYFKKDEYTFATSDAHVYMWVDOEVYINKASNS-----NKIRLEGRLY- 114
DB 966 SFTNFKKAD-----INSLNDESKRKLIEDINKTKTQLSFDLYNKKIKLERIFD 1020
OY 115 -----QIKIYOENPTEKGLDFLYTDSOKKREYVSSNLDLPELKKSSNKRKRS 168
DB 1021 KKTYGAKKMKIKKILKLEQJESKL-NSLNPKRYL-QNFSYFNKKKKEIAELEN 1076
OY 169 TSNAGPIVDRNDGCIPIPSLEVGIVDYVKRKFPLSPWINSIHKKGIITY---KSSP- 223
DB 1077 T-----LENKTKILNHY-----KGLVYTYNGSSPLK 1103
OY 224 -----EKWSTASPIYSDFE-----KYTGRIIDKNSPSPAR-----HPLVAPYPIVHD 265
DB 1104 TLSEESIQTEENYASLENFRVLSKLEGLKMDNLKRLKSLTSSGLHLHIA-----E 1156
OY 266 MENIILSKNEDOSTQNTDSERTTISKTSTSRTHSEVGNAAEVHASFIDIGSVSAGFS 325
DB 1157 LKEVYIKNNR---YTGNSPSPENNT-----DVNNKLESYKKEFLPEGDVATVVS 1200
OY 326 NSNSSTVAIDHSLSLAGERTAETGLTADTARLANANRYVNTGTPYIYVLPPT----- 381
DB 1201 ESGSDTLQSQPKKPAFTVAGAS---NTYTTTSG-NVDEVDVDYIIVPIGSEEDYDDL 1256
OY 382 -SIVLGRKQNTLATIKAKENQLSOLIAPNNYPSKNAPIALNADDDSSPTIMYNOFL 440
DB 1257 GQVYVGEAVTSYV---DNILSKI---ENEVEVLYLKPLA-----GYIR 1294
OY 441 ELEKTKQALRDTDOYVGNATATFNENGRV---RVDSGMSSEVLPJOIETARILFNKQD 497
DB 1295 SLAK-----QLENNVMTNNAVMDILANSFKRREKFAVLE-----SD 1332
OY 498 L-----NLYERILANVPSPDLETTTPKMTLKEALKAFGEFNPNGNLQYOGADITEFPFN 553
DB 1333 LPIYKOLUSSRYVAKDYKFLKKEKROKFLSSYNYI-----KSDIYDIDIN 1377
OY 554 FDOQT-----SONIKNOIAEL-----NATNYI-TVLDKI----- 581
DB 1378 FANDVLYGYTKILSEKYSPLDISIKKTYINDKROGENEKYLPFLNNIETLYKTYMDKIDLFVI 1437

```

```

OY 582 -----KLNKNANI-----LIRKRFHYDRNNINIAVGADESVYKAEHREYINS 622
DB 1438 HLEAKVLYTEKSNVEYKIKELYLTKTQDLAKDFKNNFVYIADLSTDYNNHNLTF 1497
OY 623 --STEGLLINDDOI-RKILSGYI---VEIEDEGKLEKVINDRIMPLNLISSIPDQGTFL 676
DB 1498 FLKTMVRENATVYISNLGSLNQLGMLNLSHCQVCKQCPQNRKQFRLDERECKCLL 1557
OY 677 DRKRYNKPPLPYISNRYVAVYVNTKNTINPSENG-----DTSNGIKKI 724
DB 1558 NIKQSGDKC---VENNPYCN-----EN-----NQCDDADKCTEEDSSNG-KKI 1599

RESULT 5
ID YMG7_YEAST STANDARD; PRT; 1658 AA.
AC 003661; 004988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 187.1 kDa protein in G0A1-ERG8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
PY SEQUENCE OF 1-711 FROM N.A.
RP STRAIN=S288C / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RL Walsh S.V.;
RN Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN=S288C / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A.,
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; Z49809; CAA8934.1; -.
DR EMBL; Z49939; CAA90190.1; -.
DR SGD; S0004832; YMR219W.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893P668305A757D CRC64;

Query Match 4.7%; Score 178; DB 1; Length 1658;
Best Local Similarity 18.8%; Pred. No. 0.16;
Matches 165; Conservative 139; Mismatches 286; Indels 288; Gaps 42;

OY 1 EYKQNRNLNLSSES---SSQGLGYFSDLNFOAPWVYTSYTGDLSPSSLELNPSENG 57
DB 290 EMEELDDLDVESDAKEDSQAGETHS-VDFSKYMPRTDNTKIVYIKRYSDEHVRHQ 348
OY 58 QIFQSAINS-GEIKYKSPD-----YTFATSDN---IVYKRVDOEVYINKASNK 105
DB 349 KISEGAFDFGSAIVSYVDESEDEDSQAEISTANAEVYTHNE-VEDDDDELIEDIESDS 408
OY 106 IRLERKRLYQIKYQYQRENPYFKGLDFK-----YVTSQNKKEY 145
DB 409 -----ESQASQESQSEDDPEYKMKNEKSTSELTENTS-NSDDGFAADATYKRV 459
OY 146 ISSDLQQLPEL-----KQSSNSRKRKSTAGPIVDRNDGCIPIPSLEVE 190
DB 460 EDDENDERPEKDDIIRSSLDKRFNGNNKSEYSENVLENETDAIVERENQ-IND--VA 515

```

```

QY 191 GYTVADKANKRFL---SPWISNIH- KGLITKYS-----PEKWTASD----- 231
DB 516 GYDTGKSVESDLHBSH--DNLTDLAANAQOOSRNSNCPQKEQSVSTLGHSGS 573
QY 232 -----PYSDPERVTGRIDKNSPEARHPVAAPLYHVDMENITLKNEDQ 277
DB 574 NLGSRSLDESEQJPLKDF--TGENNNLKTD- KGLISS--VEIEVKV-----SKRK 622
QY 278 STONTSETRISKNTSTSTRTSEVHGNAEVNASPFIDGSGVSGAGSNGSNSVAIDHS 337
DB 623 LDGSTEKEVLPSTDTTIN-----NSSIGNDSITYSLDA 658
QY 338 LSLAGERTVAEMGLNTA-----DTARLNINRYVTGTAPYAVL 378
DB 659 DAISENTLOVPMEIKTTPYEVVISESVSYSTSEDTNVAMPQVEK-----TSPMND- 713
QY 379 PPTSIVLGNQTLATIRAKENOLSGIAPNNYVSKNAPALMAQODSPPTM----- 434
DB 714 PENSL-----ND--DYKKHDLAKSTLA-----LAPATKKDAEVEAGVTKSCLT 758
QY 435 -----NNQOLELEKTKOLRTPDYTGAIATYFENGRRVDTGSNNSEVLPQIDETAR 490
DB 759 STSGHNIHHSKTKOVS-DLDESTEVTFTENGTG--DENKOSKNEPEVANSTDK 813
QY 491 IIFNGDLVERRIAVNSPDLFTTKPDMTLKALKIAPFEPNGNLQYOGADITEF 550
DB 814 STEDNTD--EKYFSAINTYN--VYGDSCEDIETASVVEE--NLRYCEDMDEA 862
QY 551 DENF-DOOTSQ--INKOLA-----ELNATNYTVLDKIKLANKMNTLIDRKFRYD 599
DB 863 EMSGCECKQKODDSKTOISFSTSPDNFQESNDYTESSTK-----YK 907
QY 600 RNNVAGADESVYKKAHR-EVIN-----SSTEGLLINDIDIKI 638
DB 908 VRNSDLEDESEKRLTKRAEVYDKLDESESDSYEDYADPPGCDGSGNENIYKGTAK 966
QY 639 LSGYIVIEDTGKLEVINRYDMNLISLQDQKFIIDFKYNDKPLTISPNRYKNV 998
DB 967 -----DTGIYEPENKVN-----KVHEEFTLEAVSSSNV 999
QY 699 YAVTKENTIIINSENGDSTNGIKKILIFSKKGEIGZ 736
DB 1000 -----QNKDMHDVINO--EAGANVENGE 1021

RESULT 5
YJL3_YEAST
ID YJL3_YEAST STANDARD: PRT: 1803 AA.
AC P47024.1887192.
DB 01-FEB-1996 (Rel. 33, Created)
DB 30-MAY-2000 (Rel. 39, Last sequence update)
DB 30-MAY-2000 (Rel. 39, Last annotation update)
DE Transposon Ty4 207.7 kDa hypothetical protein.
GN TY48 OR YJL113M OR J0780.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SZ88C / FY1679;
RX MEDLINE=97103775; PubMed=8948101;
RA Czaplewski C., Kordes E., Pujol A., Jauniaux J.-C.;
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT reveals 19 open reading frames including URA2 (5' end), TRK1, PAS2,
RT SPI10, GCD14, RPL1, PHO86, NCA3, ASP1, CCF7, GZF3, two tRNA genes,
RT three remnant delta elements and a Ty4 transposon.;"
RL Yeast 12.1471-1474(1996)
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC used by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch.)
CC -----
DB EMBL: Z49389; CA89409.1;
DB SDB: S0003649; YJL113M.
DB Interpro: IPR001584; Rve.
DB Interpro: IPR001878; Znf_CCHC.
DB Pfam: PF00665; Rve; 1.
DB SMART: SM00343; Znf_C2HC; 1.
DB Transposable element; Hypothetical protein.
SQ SEQUENCE 1803 AA; 207693 MW; 16DCD7284A8D52D3 CRC64;

Query Match 4.5%; Score 171; DB 1; Length 1803;
Best Local Similarity 19.7%; Pred. No. 0.4;
Matches 163; Conservative 123; Mismatches 273; Indels 270; Gaps 41;

QY 1 EVKQERKLINSESSSGILGYFSDLPQAPVWTSSTGDLSTPSELENPENQYF 60
DB 1078 QLKKTN--HETSPKGSIG--TVAKFTNNELSLKTDPSLPKLTESINHH-- 1128
QY 61 QSAIWSGPIKVKSDETTPATSDNHVTWYWDQEVYKASNSKILKEKGRLYQIKQY 120
DB 1129 -----SNDY-----STNKEV-----KF 1140
QY 121 QREN--PTKGLDFKLYWTDSONKKEVISDNLPQAPVWTSSTGDLSTPSELENPENQYF 176
DB 1141 EKENHHPPIEDIDYDM--SDQYDMESNCOGNNLKEKLYTDKNVPTDNGINVSFR- 1193
QY 177 DRNDGIDPSLEVGTVVQVKKRFTLSPIWISNHEKK-----GLTKY----- 219
DB 1194 -----LEQNTIAGSGSPVQVYKSAFLKEPSSILNKKRKRHHKNNLSYSELREDKR 1247
QY 220 -----KSPPEKWTASDPSDFEYKTRIDKNSPEARHPVAAPLYHVDMENITLTK 273
DB 1248 SKNRVKKILPDMETVSAPIKAIYINPAISKNPDKEKEYKOAY--HKEIQLMKDK 1304
QY 274 NEQSQNTDSEPTTISKN--TSTRTTSEVHGNAEVNASPFIDGSGVSGAGSNGSNSST 331
DB 1305 VPDVVKYSRSE--IPDNLVPTNTITFKRNG--YKAIYCRIDQSPPTSVITTT 1358
QY 332 VALDHS--LSLAGERT-NAETMGLTADT--ARLNINRYVTGTAPYAVL-ITSL 383
DB 1359 ESLNHHITFLMANNRNFMTLIDNHAFYAKLEF-----IYIHHDRRC 1408
QY 384 VLGKNOTLATIKAKENO-----LSQI-LAPNRYF-----SKNLAPALMAQODEFS 428
DB 1409 VAKLNKALYGLKQSPKEMNDHLKQYLNGLKDNSTTPELYOTENKMLM-IAYVDDCVT 1467
QY 429 STPTTNNTNQL-ELEKTKOLR-----LTDQVYGNATYTFENGAV----- 469
DB 1468 AASNEORLDEFINKLNFELKLTGTLIDVLDLD--ILQMDLVYKVLGTLIDTLTKSTIN 1526
QY 470 RVDTGSNMSVLPQIOTETARIENGKDLNVERIAAANP-SDFLETTAPD-----MTL 523
DB 1527 RMD--KKVVEELKIRRSIIPHSYTK-----IDPKDYLOMSEEEEROGVAKL 1573
QY 524 KEALKIATGFENPENGNYQYOGKDIYEPDFNDOOTSQNTKOLAELNATNYTVLDKITL 583
DB 1574 QQLT-----GELNIV-RHKCRDYIEFAVKAARLVNYPHERFYVYKIIQYL- 1620
QY 584 NAKMNTLIDKRFYDRN-----NIVAGADESVYKKAHR-----VI-----NS 622
DB 1621 -----VRKDGIDGHYDCKDKKVIATATASVGEYDAQSRIGVILWGMNIPVYSNK 1675
QY 623 STEGLANTINDIRKILSGY-----IYIEEDT-----EGTK----- 653
DB 1676 STNRCSVTEAEALATIEGADSETLAKYTLKEGBDNNDVMTTDSKPAIGLAKRSYQ 1735
QY 654 -----EYINDRYDMNLISLRQDKFTI-----DEKKY 681

```

DB 1736 PREKFTWIKTEIKIKIKIKITKIGKNADILTKPVASDFKRF 1784

RESULT 7
SMC3_YEAST STANDARD: PRT; 1230 AA.
ID SMC3_YEAST PRT; 1230 AA.
AC P47037;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Chromosome segregation protein SMC3 (DA-box protein SMC3).
GN SMC3 OR YJ1074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M303;
RX MEDLINE=97474309; PubMed=9335333;
RA Michaelis C., Clock R., Nasmyth K.;
RT "Cohesins: chromosomal proteins that prevent premature separation of
RT sister chromatids.";
RL Cell 91:35-43(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Rose M., Koester P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [1]
RC STRAIN=S288C;
RX SOR F.J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: Y14378; CAA74655.1;
DR EMBL: Z49349; CAA89366.1;
DR EMBL: X88831; CAA61313.1;
DR SGD: S0003610; SMC3.
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C/1.
DR Pfam: PF02463; SMC_N/1.
DR Mitosis: ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 32 39 ATP (POTENTIAL).
FT DOMAIN 172 482 COILED COIL (POTENTIAL).
FT DOMAIN 685 1041 COILED COIL (POTENTIAL).
SQ SEQUENCE 1230 AA; 14136 MW; B152D88b7780341f CRC64;

Query Match 4.5%; Score 169.5; DB 1; Length 1230;
Best Local Similarity 19.8%; Pred. No. 0.28; 355; Indels 247; Gaps 38;
Matches 182; Conservative 136; Mismatches 355; Indels 247; Gaps 38;
DB 1 EYKQENRLNLSSESSGGLGYSDANFOAPVYT--STTGGLSTPSSLEINPSNQ 58
DB 209 EMBQERKELEYENLNENKITYFTLDRLENEVINQMERLDGYN-----NTYSSQ 262
DB 59 YFO-----SAIWGPIKVKKSDYFTAFSADNHTVMWVDQEVNR-- 99

DB 263 YIQLDKREDMIDOVSKLLSI-EASLKIKNAATDLOQAKRESEISKLTNNVNTKIDVQ 321
DB 100 ---ASNSKIRILEKGLYQIK-IOYORENPTKGLDFKLYWTDSQNKREYSSDMLQPE 155
DB 322 QOIESNERQMLDSATLKEIKSIIOQRKRLSKILPVOELT---KEAMV--SLQIAS 375
DB 156 LKQSSNSKRRKSTSAQTPVPRDNDGIPDSLEVEGYTVQVKNKRFELSPWISNIE-KR 214
DB 376 LQQRDRLILKGGYAFKSKDERPWHSEIB-----ELKSS-----IONLELES 422
DB 215 GLTKYSSPEKWSASDPYSDKDEKVTGRIDKNVSPAPRHPVAAVPIVHVDENTISKN 274
DB 423 OLQMDRTSLRQYSAID-----EEIEELIDISINGPTGOL-----EDFSELIHLKQK 471
DB 275 EDQSTQNDSTFRITSKNTSRTHTSEVGNAPYASFPDGGSVASAFSNSRYAI 334
DB 472 LSESLDTRKEIMRECKLOTIVETILSDVNOR-----NVN---KSRSLAGIINKEI 525
DB 335 DHSLSLAGE-----RWAEYMG-----LNTADTALNANITYVNGT 370
DB 526 TEKLISPESTVFGTLGELIKVNDKTKCAEYIGGNSLPHIYVDTETATLNNELTRMG 585
DB 371 TAPIYVLPPTSLVGNKQTLATIKAKENLSQILPNNYTPSKNLAPALAAQDPSST 430
DB 586 GRVTF--IPLNRLSLSDVKKPPSTTQIQFTPLIKIKYEPREKA-----VKHVGKT 638
DB 431 PITMNTNPLELEKTKQLR--LDTDQV-YGNAT--YNPENGVRVYDGSNMSEVLPQI 484
DB 639 IYVKDLOGELKIAKKKLNALITLDDRADRGVLGGLDQHKRTRELSKLNESRSOH 698
DB 485 QETTALIFNGKDLNVERRIAIVNS-----DPLETKPD-MTKL 524
DB 699 KILIELDFVNEELNDIDTKIDQVGNIRKYSNDRESVLTNIEVYRISINTKNEKILLE 758
DB 525 EALK-IAGFNEPKNLOFGKDIPEFN---FPQOSTONIKNOIA-----E 568
DB 759 ESNATITLKLKLNNTNTPAQKLTNFEENLLOEFSSELSKEKERLESITWEISAANK 818
DB 569 LKNTN---IYVLRKIKLNKMLIRDRPHYRNNAV-----A 607
DB 819 LNTSALALEGTTIID--SLNLELSKILQENDELSKMSFVGAFPLGDELKEQLQLE 876
DB 608 DESVYEAHREYVNSR-----EELL--NIDKIRKI 638
DB 877 KESEVKQENNAVLELGVQRETESLAEETNNKKLLEKANNQQLKLDLNFQKSYEKT 936
DB 639 LSGYVLEIEDYEGSLK-----VINDRTM-----ISSLRDQK 673
DB 937 MIKTTLVTRREELQRIREIGLLPEDALVNDSDITSDOLLQRLNDMTETISGLKYNK 966
DB 674 -FEIDFKKYNDK-----LPYIISNPYKVNVAVTEKNTIINPSNG 714
DB 997 RAEENKFKFENRRDLAERASELDESKDSIODLIVALKQOKVNAV---DSTFOKVSBNF 1052
DB 715 DT-----STNGIKKILIFSK 729
DB 1053 EAVFERLVPRGTATLTHRK 1072
RESULT 8
WAPA_BACSU STANDARD: PRT; 2334 AA.
ID WAPA_BACSU
AC 007833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G.
OS Bacillus subtilis.
OC Bacillus; Firmicutes; Bacillus/Clostridium group.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;

[1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168:
 RX MEDLINE-93302506; PubMed-8316082;
 RA Foster S.J.;
 RT "Molecular analysis of three major wall-associated proteins of a
 RT Bacillus subtilis 168: evidence for processing of the product of a
 RT gene encoding a 258 kDa precursor two-domain ligand-binding
 RT protein."
 RL Mol. Microbiol. 8:299-310(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / BGSCLAI;
 RX MEDLINE-95219088; PubMed-7704263;
 RA Yoshida K.-I., Sano H., Seki S., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 RT genome containing the hut and wapa loci."
 RL Microbiology 141:337-343(1995).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / BGSCLAI;
 RX MEDLINE-97124196; PubMed-8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 RA Miya Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the hic and cel loci, and creation of a 177 kb contig
 RT Microbiology 142:313-323(1996).
 CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 CC MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 CC INTO THE MEDIUM.
 CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 CC MOTIF REPEATED 31 TIMES.
 CC -1- SIMILARITY: THE REPEAT IN E. COLI RIB GROUP OF PROTEINS (RHS-A-D).
 CC SIMILARITY TO THE REPEAT IN E. COLI RIB GROUP OF PROTEINS (RHS-A-D).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 DR EMBL: I05634; AAA22883.1; -
 DR EMBL: D31856; BAA0656.1; -
 DR EMBL: D29985; BAA0626.1; -
 DR EMBL: D83026; BAA1168.1; -
 DR EMBL: Z89124; CAB1593.1; -
 DR PIR: S32920; S32920.
 DR Subtilisin; BGI0797; wapa.
 DR InterPro: IPR003305; CBD_6.
 DR Cell wall; Repeat; Signal; Complete proteome.
 KW STGNAL 1 28 OR 32 (POTENTIAL).
 FT DOMAIN -29 2334 WALL-ASSOCIATED PROTEIN.
 FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDDEM REPEATS.
 FT REPEAT 504 605 1-1.
 FT REPEAT 636 736 1-2.
 FT REPEAT 769 869 1-3.
 FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDDEM REPEATS OF
 FT REPEAT 1021 1040 X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
 FT REPEAT 1042 1061 2-1.
 FT REPEAT 1063 1082 2-2.
 FT REPEAT 1083 1102 2-3.
 FT REPEAT 1109 1128 2-4.
 FT REPEAT 1129 1148 2-5.
 FT REPEAT 1150 1169 2-6.
 FT REPEAT 1174 1193 2-7.
 FT REPEAT 1174 1193 2-8.

FT REPEAT 1199 1218 2-9.
 FT REPEAT 1219 1238 2-10.
 FT REPEAT 1646 1665 2-11.
 FT REPEAT 1667 1686 2-12.
 FT REPEAT 1690 1709 2-13.
 FT REPEAT 1711 1730 2-14.
 FT REPEAT 1732 1751 2-15.
 FT REPEAT 1753 1772 2-16.
 FT REPEAT 1795 1814 2-17.
 FT REPEAT 1820 1839 2-18.
 FT REPEAT 1840 1859 2-19.
 FT REPEAT 1861 1880 2-20.
 FT REPEAT 1887 1906 2-21.
 FT REPEAT 1908 1927 2-22.
 FT REPEAT 1929 1948 2-23.
 FT REPEAT 1969 1982 2-24 (APPROXIMATE).
 FT REPEAT 1983 2002 2-25.
 FT REPEAT 2008 2027 2-26.
 FT REPEAT 2028 2047 2-27.
 FT REPEAT 2051 2070 2-28.
 FT REPEAT 2071 2090 2-29.
 FT REPEAT 2093 2112 2-30.
 FT REPEAT 2120 2139 2-31.
 SQ SEQUENCE 2334 AA; 258329 MW; B75138CDD278BAA3 CRC64;
 Query Match 4.3%; Score 163.5; DB 1; Length 2334;
 Best Local Similarity 17.8%; Pred. No. 1.3; DB 1; Length 2334;
 Matches 170; Conservative 123; Mismatches 338; Indels 325; Gaps 40;
 1 EVKQENRLNLESSSSQGLGYFDLNFQA-----PRAVTSSTTGGDLSIPSSLEINIP 54
 269 EYERSDKVSYLKENEGILHLADENMLKDERVYPVSTDPSTLSVSDTFVMSAYP 328
 55 SENQYFQSAINSGFK--VKSDEYTPATSDNHYTM----- 89
 329 TTVNSASSQKDAHLKAYLKTGYD-KTGTNYAMKPFNLKTLNMTVTRATLKYVA 387
 90 -----WDDQEVINK-----ASNSKIR-LNGR-LYQIRIYOR 122
 388 HAYYGRKATGLMD--TVNSYDAAKAVYNTKPSKNTGRADYKQVMSYDTAAVKS 444
 123 ENTETGDLDFLYWTDSQNK---KEVTSNDLQPELKQSSNKKKRRSTSGPVP--- 176
 445 WNSGAGNNGFRLH-TGNGKEVYKTLIS-----SANSANKPYIEVYTIPIRGN 491
 177 -----DRNQGIDP---SLF-VGGYVDYKNEKTFISPMI-SNHEKGLTKYSSP 223
 492 TPTIKAYINGDSIGYDIDSMKVEGAKGYVNIYNCKEYQALSNQV-----TMSYKG 545
 224 EK-WSTASDPYSDPEYVGRID-KVNSPEARHPLVAATPIYVHNMENILSKNE----- 275
 546 KTIWPTSAEIASKRYKL--HLDGKDGLALDP---SPVTKNSGSIYATSKNWTGYSA 599
 276 --DOSTONTDETRTISKNTSTSRTHGSEVHNAEYHASFPGDGSVAG-----FS 325
 600 IFQOEGGAMSPAPVPIPNVGAQAPSAKYNNGNA--TGYPDLSMKVAGATGYKVQVFN 658
 326 NSNSSTVAIDH--SLSLAGERTWAEIWTGLTADTA-----RLNANTRYVNGTA 372
 659 GKGFETLDLIGNQTSWTTKGKTWPTSAIKAGYALHDKSGAEPLIPNGEPTVNNAGCD 718
 373 PIYVNLPTTSVLCKNOTLATIKAKENQLSQIAPN----- 408
 719 GARRNYSFKIAYNKGDEAIASPAATPALDIPARNVNTGLYVNIKRSQTYVNLWEK 778
 409 -----NYPKRLNAPLANODPSSPTITMYNOFLEKTR----- 446
 779 VQNAKGYVNIYNKEVQSFVDGADAHY---TONKNIMPTSELKASGYKLHTDGGK 834
 447 QLTLDTDQVGN-----IATYN--FENGRRVVDGSMWS 478
 835 ELALDPSFYNNANGNYKGNKRYSTLVAYDANGERTIPALPFPNTHBEAEELGTBEYWS 894

```

OY 479 EV-LP--OIOETARIENGKDLN-----VERRIAANPSDPL-----ETT- 517
DB 895 IIPBGGONATNVMEEDSLDGGPGLDLSSTWSSDHLFGGQWYADAEATSV 954
OY 518 -----KPKTLKALKIATG-FNEPNCLOYOGKDIFEDFEDQOOSNIKQL 566
DB 955 ISFDGAMTIDEDATTHREFKRAADGTOPPLGYALELETADOFILTKDQONAVENKRG 1014
OY 567 AEL-----NATNIYVLDKIKLAKMNTLIDKRFHYDRN-----NIAVGA 607
DB 1015 GKLOKVVGDHNNNAVYVYNDKNOALTAIDASGRKLFFYDENGHTVSTIGPKNKVTYTSY 1074
OY 608 DESVKE-----AHR--EVINSSTE 625
DB 1075 ENLLKRYDNDGIVTSYDSEGRLYKOYSANSTAKPFTETGYSGHLEKKAINKAKE 1134
OY 626 GLLNTDKDI-----RKILSGY-----IYIEETGELKEVINDRYDLMI 665
DB 1135 TVYYSYDADKRTLLMOPNGKRVGYGVNAGNPLOYIDAGSLKITNTKRYGNNV 1190

RESULT 9
P115_MYCHR STANDARD: PRT: 979 AA.
AC P41508:
DB 01-NOV-1995 (Rel. 32, Created)
DB 01-FEB-1996 (Rel. 33, Last sequence update)
DB 15-JUL-1996 (Rel. 36, Last annotation update)
DB P115 protein.
DB Mycoplasma hyorhinis.
DB Bacteria: Firmicutes; Bacillus/Clostridium group; Mollicutes;
DB Mycoplasmaetaceae; Mycoplasma.
DB NCBI_Taxid-2100;
DB 11
DB SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
DB MEDLINE-91138990; Pubmed-1825306;
DB Notariicola S.M., McIntosh M.A., Wise K.S.;
DB "A Mycoplasma hyorhinis protein with sequence similarities to
DB RT nucleotide-binding enzymes.";
DB RT Gene 97:77-85(1991).
DB CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
DB CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
DB CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
DB CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SIMILAR TO OTHER MYCOPLASMA
DB P115.
DB -----
DB This SWISS-PROT entry is copyright. It is produced through a collaboration
DB between the Swiss Institute of Bioinformatics and the EMBL Outstation-
DB the European Bioinformatics Institute. There are no restrictions on its
DB use by non-profit institutions as long as its content is in no way
DB modified and this statement is not removed. Usage by and for commercial
DB entities requires a license agreement (See http://www.isb-sib.ch/announce/
DB or send an email to license@sib-sib.ch).
DB -----
DB EMBL: M34956; AAA25423.1;
DB PIR: J00894; J00894.
DB InterPro: IPR003439; ABC_transport.
DB InterPro: IPR001687; ATP_GTP_A.
DB InterPro: IPR003405; SMC_C.
DB InterPro: IPR003395; SMC_N.
DB Pfam: PF02483; SMC_C.1.
DB Pfam: PF02463; SMC_N.1.
DB ATP-binding; Coiled coil.
DB NP_BIND 32 39
DB FT DOMAIN 169 224 COILED COIL (POTENTIAL).
DB FT DOMAIN 231 400 COILED COIL (POTENTIAL).
DB FT DOMAIN 569 821 COILED COIL (POTENTIAL).
DB FT DOMAIN 884 912 ALA/ASP-RICH (DA-BOX).
DB SO SEQUENCE 979 AA; 110566 MW; 30D51C5656280F4 CRC64.

```

Query Match 4.3%; Score 162.5; DB 1; Length 979;

```

Best Local Similarity 18.9%; Pred No. 0.47.
Matches 175; Conservative 133; Mismatches 341; Indels 277; Gaps 40.
OY 13 ESSSGGLGYFSDLPNOAPMY-----TSSTGDLSTPSSLENIDSE 56
DB 51 EDSAGQLGMDVDVIFASKTVPKOEKAMVLTFFNEDALIEKQITSLRLRGQT 110
OY 57 NOYFOS-----AIVSGFTKVKSDYFTFSADNHVYVWVDOE--VINKSN 102
DB 111 NEFYFDDQPVPRKDKMLAVESG---IKSSLAIIISGTSIELANALPEKRAVIEEAG 167
OY 103 SNRIELGKRLYQIKIOYORENPTEKGLDEKLYMTDSQNKKEVSSDNLQBELKQSSN 162
DB 168 TSRYKTKDEEAKKLIR-----TNDALD-KLOGAIKELEOVAS_LDK--QASKATYLEK 219
OY 163 SRKRRTSAGPLVDRD-----NDGIPDSLEVGTYTVDK-NKHFLEPSWISNHEKKG 215
DB 220 SKALSEVGEGLIVNDLNFENKKNLNTSLLEVDQQRNDLELNQITRESSISQYH---- 275
OY 216 LTKYKSPKKSSTADPSYDPEKVTGRD--KNVSP-----EARRHVAAYPIYHV 264
DB 276 --FKTEVE-----SSIOEITSKDNLKNALSEINLOEARIEBRRLIISGEIV-V 322
OY 265 DMENIT--LSKN-EDOSTORTSEPTITSKNTSTSTHT----- 300
DB 323 DQKTKIEIKQVESLKIQINASKORETELDQULFLAKANSK_KVENDINKEIGVLE 382
OY 301 --SEVGNAEVHASFEDIGSVSACFSNSSTVAIDHSLSIAGKRWAE--TGLMTA 355
DB 383 KKSAAANINILKQOFENKSFLSKIGITIKONSELPDGYIGLASELFKRVSEFSIATITV 442
OY 356 DVARLN-----ANIRVNTGTAPIYVNLPTTSL-----VLG 386
DB 443 LGNALQIYMTSSEDVLOADIFLKRNLGSKATFPLTSTIKEREVEDHLVLKQGRFLG 502
OY 387 KNOTLATIKAKENO-----LSOILAPNNVYPSKULAPI-----ALN 422
DB 503 VARELEFPQGNLPGFLLGNLVVDVANDANRIKILDKYTVSLEGDLEPRGGYIT 562
OY 423 AODDFSPPIPMNNOFLEKTKQLRLDTQ-----VYCONLATVFNENGRVAVD 472
DB 563 GSGKLEKRSI-LNIDIK-EHTNKLKAEQDHDKIKQITINELIETVNSTIQVATIE 620
OY 473 TGS-----NMSEVLPIQIE--TTRILFNGKQINVERIAAV----- 508
DB 621 ANSINKNLINIELNMLKINSELPFEQDQESLNTSEFSEKLN-IEQVISTLTIELN 679
OY 509 NPSDPL-----ETTKPDMTLK-----EALKIAGFNEPNNGIYQCK 545
DB 680 SKKORLTNLSQKRGKTEKQELDKLKLNTQHSSTEQNRAKF--VEDONKRUSEHYK 739
OY 546 DITE-----FDENFDQOTS--QNIKNQALNANNTVYVLDKIK_NAKMNTLIDRRF 596
DB 740 LLELAASEQYSLDIDIOAHHPVDSIKKELEKELGNLLEAITEFEVENOR---YQERKO 795
OY 597 HYDRNIAVGADESVEVKAHREVINSSTEGILLNIRKD-----IKTLISGYIYELEDTEG 651
DB 796 YIELTATKASIEEALISDLKILINKTE--IYNLVNNEFNMYVQKFKGGAKEIHFT-- 851
OY 652 LKEVINDRYDLMUNTS--SLRODGKTFIDFKYND-----KIPLEVI- 689
DB 852 -----DKNDILNSGVEISAGPQKTKIKNLILFSGEKAIIAISILFAIIKARPILCIL 905
OY 690 -----SNPNKKNVAVVAKENT 706
DB 906 DEVEAALDESNAVIRYVEFLKLAKENT 931

RESULT 10
YD86_SCHPO STANDARD: PRT: 1957 AA.
AC Q10411:
DB 01-OCT-1996 (Rel. 34, Created)

```

01-OCT-1996 (Rel. 34, last sequence update)
 01-OCT-1996 (Rel. 34, last annotation update)
 DB Hypothetical 222.8 kDa protein Clp3.06c in chromosome I.
 GM SPAC1F3.06c.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 RN NCBI
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Connor R., Church C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: 270690; CA94624.1;
 DR Hypothetical protein.
 DR Hypothetical protein.
 DR SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 4.38; Score 162; DB 1; Length 1957;
 Best Local Similarity 20.4%; Pred. No. 1.3;
 Matches 177; Conservative 132; Mismatches 328; Indels 232; Gaps 40;

0Y 2 VKQENLLNNESE-----SSSQGLGYTFSDLNFQ--APWVYTSSTTGDLSPSSSE 49
 DB 129 VFOKSNLNEKOVRSKLALEHENGILSLQLSSSKKMKNTSVTLTSEEDVSYFPKK 188
 0Y 50 LENISENQYQSAIMS--GFIKY-----KSDDEYFATSDNHVVMWVDOEYFNAS 101
 DB 189 LTNMESNSAKOSAVDYLSROLITVEKLDKKEDEYKIKEDVSISAKASAEQASNSL 248
 0Y 102 NSKRIELEK--GLVQIKIYQRENPT--KGLDFKLYTDSQNKKEVSSDNQJPE 155
 DB 249 RGQRELEKILVSNKTVSTLQNTENSILRAVCTQLEK-----EKCALNEDSKLEE 302
 0Y 156 LKQKSN-----SRKKRSTASGTPVPRDNDGIPSLVEGVYVADYKRR----- 200
 DB 303 LKNNVANSDAIVHKDLIEDLSTISFPD-----LKSEHDLISIKNEKLEKLNRTI 356
 0Y 201 -TFLSPWISNHEKKGLTKYSS--PEKSTASDPYDEPK-----YTRGD----- 244
 DB 357 GSLKDSRTNSOLEEWEVLEKESNTIHSQLDPSKSLSPQENKSLKGSIDERYNNLS 416
 0Y 245 -----KWS--PEARHPLVA-----YPIVHDM----- 266
 DB 417 SKDKMYKOVSSOLEEASSLAHATGKLAELNSERDFONKKIKDEFEKQDLRACLNSSN 476
 0Y 267 --ENILSKNEDOSTOND--SETRTISKNTST--RTTSEVHNAEVAHASF 315
 DB 477 ELKEKSLIDKRDDELNLRQIKQKQVSESSQSLQSLQDLINKKHEVYSQUNE 536
 0Y 316 IGGSVASGASNSNS--STVAIDHSILAGEPTVAETMGLNTADPARANANIRVNTG 370
 DB 537 LKGELOTEINSEHLSQSLTLAAKKAVALTNNEISEK--NSIQTL--CNA----- 585
 0Y 371 TATIVAVLPTTSLVIGKNO--TLATIKAKENQSLIAPANNYPKRNAPALNAODD 426
 DB 586 ---FOEKILAKSVMLAKENQNSISLDTSPKKNMESQELNNHQTIKOL-----KD 634
 0Y 427 FSGSTPTTMYNOFLEKTKQKRLDTPDYGNATAYPENGAVYVDGSMKSVLPQOE 486
 DB 635 TSS-----KQOOLQERANFPQKESITLSDANDLTK-----LTKLEE 672
 0Y 487 TTRATIFNGKDLNIVERRIAAVNPSPDLETTRKPD--TLKEALKI-----AFGENEPNGML 541

DB 673 SNKSLIKQOEIVSLKNTI-----OYLKEDLEKSEALRFSKLEKLNREYIDNL- 722
 0Y 542 YQKQDIT-EFDNFEDQSTQNIKQALAEATNATYTVLDKIKLAKKNWILIDKRFHYDR 600
 DB 723 -GKHEHLEAQRDLNLSLSDAKNTNAILSESLTSSSDVRRLVANVETLTQDSK----- 776
 0Y 601 NNIAVGADESIVKAEAREVINSSTEGILLNIDKIR-----KILSGYIVEI-----DT 649
 DB 777 -----AMKSFSLVNSYOS--ISNLYHELDDHVMQOQNNTLSESESKLTDC 824
 0Y 650 EGIKE--VINDRDMILNITSLQDCKTFIDPFKYNKLPYISNPYVANYVAATKENT 706
 DB 825 ENLTQNNLTLDVQKLMKHQNEQSV--SELKEVNGSLSDLKNLRSSLVN-AISDNQ 882
 0Y 707 IINP-----SENGDT-----STNGIKI 724
 DB 883 ILTQLAELSKNVDLSLEQSAQNSGLKSL 911

RESULT 11

USOL YEAST STANDARD; PRT; 1790 AA.
 AC P2386;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 GN Intracellular protein transport protein USOL.
 OS USOL OR INT1 OR YD058W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN NCBI
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180-1A;
 RC MEDLINE=91185402; PubMed=2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yamaoka M.;
 RT "A cytoskeleton-related gene, usol, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae."
 RL J. Cell Biol. 113:245-260(1991).
 RP SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,
 RA Kendrick K.E.;
 RL Submitted (Feb-1993) to the EMBL/GenBank/DBJ databases.
 RN NCBI
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL. SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YB1047C FAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X54378; CA38253.1;
 DR EMBL: U03188; AB000143.1;
 DR EMBL: U53688; AB06659.1;
 DR PTR: A36455; A36455.

DR HSSP: P80220; LDIP.
DR SGD: S0002216; USOL.
DR InterPro: IPR002017; Spectrin.
KW Transprot: Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEBDDE (IN REF. 2).
FT CONFLICT 1790 1790 D -> DEBDDE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 4.3%; Score 161; DB 1; Length 1790;
Best Local Similarity 18.5%; Pred. No. 1.2;

Matches 144; Conservative 151; Mismatches 263; Indels 220; Gaps 37;

QY 1 EYKQENRLNLSSESSQGLGYFSDLNFOAPVWYTSSTGD-----LSIPSELE- 51
DB 1036 QYKEE--ITSKDSKMD---EYESQISLTKERLKTATANDENVKNISETKTRLEELA 1089
QY 52 -----NIPSENOYFQSAIWSGFIKVKKSDY-----TFATSAOHVMMWVDDQEV 96
DB 1090 ELAAVKNKLNLETKLETKETSEKAKKEVENEHLKKEKIQLEKATETKQQLMSLKANES 1149
QY 97 INKANSNKITLEGRLYOIKI-----OY-----OREPREKLDKRLW- 136
DB 1150 LKEHEBDLAQAKK---YEOJANKEROYNELSQUNDLITSTQENESIKKKNLELGE 1206
QY 137 -----TSONKKEVYSSNDLQELPKOKSSNRKRTSGPVPRDNDQIPSL 187
DB 1207 VKKMSSTSEBQNSKSELDALNDIKELKKK-----ETNEASLDSI 1250
QY 188 E-VEGYIVDV- -NKRFLSPWISNIHEKKGGLTKYKSPKMWSTADSPYSPKVTGRI 243
DB 1251 KVSSEYTKIKELDDECNREKREVESELD-----KIKASEKNSKYLELQEKSEKIKEL 1305
QY 244 DKVNSPEARHPLVAAPVIVHVDENII-LSKNEDOSTONTDSETRTISKNTSTSRTHSE 302
DB 1306 DAKTTE-----LKIQLEKTNLSKAKES-----ESELRLKTKTSSEERKNAEE 1349
QY 303 ---VHGAAYHASFEDIGSGVSAGFSNSSTVAIDHSLSLAGERTWAEGLTADTA 358
DB 1350 QLEKIKNEIQIKNAQEP---KERRLNBSSTIITQEV---EK-----INTLE- 1391
QY 359 RLNAIRVYVNGTAPIVNLPPTSLVLAGKNOTLATIKAE-----NOLSOILAPNNYPS 413
DB 1392 -----DELIRJONEN---ELKAKELIDNTRSELEKVSISDELLE 1427
QY 414 KNLAPIALNADDPSS--TPTMANNQLEKTEKQULRDIDVOYGAIATVNGRGRV 471
DB 1428 EKQNTIK-SLQDELSTYKDKITIRNDEKLSIRDNKRLDSEK-----EDLRQAQ 1476
QY 472 DTSQMSSEVLOIQTAR--IIRNGDL---NLVGRILAVNSDPLET-TKPRMT 522
DB 1477 ESKAKYEBLKKEBSSKKEKELKESKEMAKKLESTISNETELKS--MEIKRSPEK 1534
QY 523 LKALAKIAPGNEPNGLQYOKDITPEDNPOOTSINKOL-----ALNAT--N 573
DB 1535 LEOSKSA---EDYIKNLQHEKSDLS-RIKSEKQIEKLSKRITANSGLSELTVAOE 1590
QY 574 IYTVLQIKLAKAKNLLIDKRFHYDRNNIATVAGADESVKHAHEVINSSTEGILLNDK 633
DB 1591 LNAOERIKIRINAEENTVLSK-----LEDIERELDKQAE---IKNSQ 1630

QY 634 DIRKILSGYIEED-----TEGLKEYINDRDMINISLROCKTFIDFKYND 683
DB 1631 EKEKELTSRUKELDELSTOQAKQKSEERRAEVRKFOVEKSQLCKAMLETKRYND 1688
RESULT 12
BXD_CLOBO STANDARD: PRT; 1276 AA.
AC P19321;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type D precursor (EC 3.4.24.63) (BoNT/D)
GN BoNTD
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BVD/-3;
RX MEDLINE-91016853; PubMed-2216736;
RA Sunagawa H., Ohya T., Kurazono H., Popoff M.R., Eklund M.W., Yanaguchi G.,
RA Binz T., Kuzajono H., Popoff M.R., Eklund M.W., Yanaguchi G.,
RA Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;
RT "Nucleotide sequence of the gene encoding Clostridium botulinum
RT neurotoxin type D.";
RL Nucleic Acids Res. 18:5556-5556(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CB16;
RX MEDLINE-93042276; PubMed-1420572;
RA Sunagawa H., Ohya T., Watanabe T., Inoue K.;
RT "The complete amino acid sequence of the Clostridium botulinum type D
RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT phase d-16 phi genome.";
RL J. Vet. Med. Sci. 54:905-913(1992).
RN [3]
RP PARTIAL SEQUENCE.
RC STRAIN-D-5A, AND D-1873;
RX MEDLINE-89359741; PubMed-2668193;
RA Koriishi K., Syuto B., Kubo S., Oguma K.;
RT "Molecular diversity of neurotoxins from Clostridium botulinum type D
RT strains.";
RL Infect. Immun. 57:2886-2891(1989).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE-94230352; PubMed-8175689;
RA Yamasaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.,
RA Roques B., Eykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinum neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -I- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 60-KDa-1-LEU-61 BOND OF
CC SYNAPTOBREVEIN-1 AND -2.
CC -I- CATALYTIC ACTIVITY: limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO
CC detected action on small molecule substrates.
CC -I- SUBUNIT: DISULFIDE-LINKED HEMERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -I- SUBCELLULAR LOCATION: secreted.
CC -I- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -I- BOTULINUMS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC BACTERIOPHAGE.

```

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54254; CA38175.1;
DR EMBL: S49407; AAB24244.1;
DR PIR: S11455; S11455.
DR HSSP: P10845; 3BPA.
DR MEROPS: M27.002;
DR InterPro: IPR000395; Bontoxilysin.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOXILYSIN.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydropase; Metalloprotease; ZINC.
DR CHAIN 1 442 BOTULINUM NEUROTOXIN D, LIGHT CHAIN.
DR METAL 443 1276 BOTULINUM NEUROTOXIN D, HEAVY CHAIN.
DR ACT_SITE 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
DR METAL 230 230 ZINC (CATALYTIC) (BY SIMILARITY).
DR METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
DR DISULFID 437 450 INTERCHAIN (PROBABLE).
DR VARIANT 15 16 ND -> PV (IN STRAIN D-SA).
DR VARIANT 17 18 ND -> LO (IN STRAIN D-1873).
DR VARIANT 452 452 R -> Q (IN STRAIN D-SA).
DR VARIANT 457 457 R -> F (IN STRAIN D-1873).
DR VARIANT 462 462 A -> D (IN STRAIN D-1873).
DR VARIANT 489 489 K -> N (IN STRAIN CB16).
DR VARIANT 644 644 N -> K (IN STRAIN CB16).
DR VARIANT 1122 1122 Q -> R (IN STRAIN CB16).
DR SEQUENCE 1276 AA; 146871 MW; C1RC50F46C8233E2 CRC64;

Query Match 4.28; Score 160; DB 1; Length 1276;
Best Local Similarity 20.68; Pred. No. 0.89;
Matches 164; Conservative 117; Mismatches 291; Indels 224; Gaps 41;
QY 44 SISSSELEIPEENQYFQSLMGFIKKKSEDTYATSDNHNTMWDQEVYKNSNS 103
DB 310 TTPSWISNI--DKY--KKIFSEKINEDKNGNVEVNIIDFNSLISDLTVNASEVYS 364
QY 104 NKIRLEK-----GLYKIKIYOQRENPTEKGLDFRLYWDSONKKEV 145
DB 365 SQYVKNRTHYFSRAHLYEPANILDNITTRDGF--NLTNKGFNIE----- 409
QY 146 ISSDNLOI--PELKOKSSNSRRKSTAGRPVDPHDNDGIPDSLEGEYTVVKNKRTFLS 204
DB 410 NSGGINTEHPALOKLSESEYVDLFKVLRLTKNSRP-----DSTCYKVNNTL-- 458
QY 205 PWISN-----THEKKGLTKYKSPKEMSTASDPYS--DFEKVGRIDKNVSPKARHPLV 256
DB 459 PYVADKSDISQIEFNKIITD--ETVQVNY--SDKFSLDESILGQVY--INBEIADVL- 511
QY 257 AAYPIVAVMEHIIISKNE-----DOSTGNTD-----SETITTSKRTS--TSRPHTESEV 303
DB 512 -----LPIVMMEDLNPGEELVIFDITFYVDLYLSYYSLEQKLSNNVENIMILTSVSEA 567
QY 304 HGNAEVHASF-----DIGSVAGFSNSNSSTYALDHSISLAGRIMWATM----- 350
DB 568 LGSYNSKIYFLLPSLAEKYNKGVQAGFLPMANNEVEDPTITNMKDLIDKISVSYIIPY 627
QY 351 -----GLNNAFTA--RLNANIRVYNTG-----PAPYVNLPTTSIVGKNQOTLATIK 395
DB 628 IGRPLNIGNSALRGFNQAFATGVAFLLEGFPFTPLPAGVTFYSSIDGRKIKIKITE 687
QY 396 -----AKENQSLQILANMN-----YPSKNLAPIALNADDDSSSPIPTM 434

```

```

DB 688 MLEBQVAKRWKDYOMNYSNMLSRITTOFNHINYOATDS-----LSIQADAIKATIDL 740
QY 435 NINQLELEK-----TKQLNLDIDVOYNT-----ATYNEPGRVAVDGSN 476
DB 741 EKKRKSQSDKNIKISOVENKAKNSLDVAKISEAMNINKFIRECSYTLTFPN----- 790
QY 477 WSEVLPJOJETTARIIIFNGKD--INTL-----VERRIAAVNSDPLETKRPDMTL 523
DB 791 -----MLPKVIDELNKFILRTKTELNLIDSHNIIYGEVDRILKAKVNES--FENRMP----- 841
QY 524 KEALIAFG--FNEPNCNLOYOGKIDTEFDEPFOOTS-----ONIKQALAEIATNI--YTV 577
DB 842 -----FNISYTNSSL--KDIINEYFNSINDSKILSLONKKNALVDTSGVNAEVRV 891
QY 578 LDKIKLNKNNIILRPKRFHYDRNNIYVGADESVEKEAREVINSSTREGILINDKDIRK 637
DB 892 GDNVQNLN--FIYNDFKLSSGKIIYVNNNIIYSA--IYENSVEFWIKISNDLTN 945
QY 638 IISGYIV--ETEDTEGLKEVINDRMDNITSIRDSKFTFDEKYNDKPLPY---ISNP 692
DB 946 SHRETTIINSIEONSCMKLICNG-----NIEMILQVY-----KKKSLIFPYSLSLST 996
QY 693 NKVANYAVTKENTII 708
DB 997 GYTNKFEVYTTNNIM 1012

RESULT 13
FAT1_SCHPO STANDARD; PRT; 1385 AA.
AC 013735; G9UT0;
DR 15-JUL-1998 (Rel. 36, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Actin interacting protein 3 homolog.
GN FAT1 OR SPAC15A10.15 OR SPAC15E1.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxId=4896;
RN [1]
RP SEQUENCE OF 1-1033 FROM N.A.
RC STRAIN-972;
RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1023-1385 FROM N.A.
RC STRAIN-972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Lyne M.H.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION, AND GENE NAME.
RX MEDLINE=20143585; PubMed=10679021;
RA Jin H., Amberg D.C.;
RT "The secretory pathway mediates localization of the cell polarity
RT regulator Alp3p/Bud6p."
RT Mol. Biol. Cell 11:647-661(2000).
CC -1- FUNCTION: INVOLVED IN THE ORGANIZATION AND/OR FUNCTION OF THE
CC ACTIN CYTOSKELETON.
CC -1- SIMILARITY: TO YEAST BUD6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z97208; CAB01112.1;
DR EMBL: AL109770; CAB52420.1;

```



```

OY 167 R-----STAGPTVPDRNDGIPDSLEVEGYVDVANKRTEFLSP-----WISNIHEKG 215
DB 248 RRRKSSSGNDNDINNDDDDANEDDSATL-----TKATITSEFLSDIWLIS-----ES 299
OY 216 LTRKSSPEK WSTASDPISDEKYTGKIDKNSPAPRPAAPVAVYVDMENITLSKN 214
DB 300 LVNOSTYLNKIMTITNOP-----NENSS-NSPLV---PLEFKINOMLTLR- 341
OY 275 EDQ-----STONTDSEF-----RTISKNT 293
DB 342 QDDYLNFTIRSEKSVDMKLVHVDISLMDPEFKITSDTSEPGIIELEYDONISKCL 401
OY 294 STSRTHSEVGNNAEVHASFIDIGSVAGFSNSNSVVAI-DHSLS--LAGERTWETM 350
DB 402 SPNNKESP-----ADTQACVGFALALAIASANAPLDSDISGPNISLTPROJASPESTALV 457
OY 351 GLWTAOTALNANRYVNGTAPYVNLPTTSVLGKNQOTLATIKAKENOLSOILAPNNY 410
DB 458 DI-----MINORGAALNT-----TVSYVI---ELIRKNSQYQVNLTLTITKT 498
OY 411 YPSKMLAPAL-----NAODFSSTPTMKNYNOFLEETKOLQTLTDVYGNIAITY 462
DB 499 HPSNDPITLYLGRKFSNHLSDP-----FQIITDENDANITLHNOLEHEKREPL 550
OY 463 NENGKRVYDGSWSEFVLPOIETTAIIFNGKDLNVERIAVNP--SDPLETKPD 520
DB 551 GFERFVY-----VELIAELLHCSNGLMNSKRAERIAARRRORVRSOLSHLDALND 602
OY 521 MTL--REALKIAFG-----FNEPNCNLOYQKDI-----TEPDFNF--QOOTSQ 560
DB 603 LSTEEEOQLKRRHSPTRDTHDKNNKIDNNDNDESDYDEIDESFEIPIYNNKO 662
OY 561 NIKQALAEINNTNITYVLKIKLNKAKNMLIRKRFHYDRNNIAVAGD-----ESVYEA 615
DB 663 NIKLRDPVTGCT-----YKSNMYMRD--CFQNNLEFLTPMNNMHWAVYDI 708
OY 616 HREVIN-----STEGHLLNID--KDIKILSGIYEI--DTEGLKIVYDRDMLNI 665
DB 709 IQDIFNGRMDSYNSPLVSLFLNLSKSYQFWTDIYVSEKGDVRSRSPYIDBPNEFKI 768
OY 666 SS---LR--QDGKTFIDFKYNN-----DKLPLYTSN-----691
DB 769 TTFEILRGQDSYFYELKRNMLGTMGHVILAEVYVNSKIAVYDVISKYSNLOTDEM 828
OY 692 PNKVVAVYVTK-----ENTINPSENGCTSYN 719
DB 829 QYISEELNEMTMMYKSKILGGSYIDGNCNIIIPOLDNPTIVLP--NGDASN 880

RESULT 15
CAGA_HELPJ STANDARD: PRT: 1167 AA.
ID CAGA_HELPJ 09ZLN1;
DT 16-OCY-2001 (Rel. 40, Created)
DT 16-OCY-2001 (Rel. 40, Last sequence update)
DT 16-OCY-2001 (Rel. 40, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein)
DE (CAG pathogenecity island protein 26).
DE CAGA OR CAI OR CAG26 OR JHP0495.
OS Helicobacter pylori 399 (Campylobacter pylori 399).
OC Bacteria: Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling I.-S.L., Moll P.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deGange B.L., Carmel G.,
RA Tummino P.J., Caruso A., Ulla-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Heiberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust J.O.;
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.;

```

```

RL Nature 397:176-180(1999).
CC -!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: AE001483; AAD0673.1; -.
CC InterPro: IPR004355; IYSEC_CAGA.
CC PRINTS: PR01553; TYPE4SSCAGA.
CC Antigen: Complete proteome.
CC DOMAIN 246 249 POLY-THR.
CC DOMAIN 882 889 POLY-ASN.
CC SEQUENCE 1167 AA; 129729 MW; FD5E6B81CEBD0F2 CRC64;

Query Match 4.18; Score 156.5; DB 1; Length 1167;
Best Local Similarity 20.68; Pred. No. 1.2;
Matches 168; Conservative 136; Mismatches 29; Indels 213; Gaps 45;

OY 3 KOENELNSESSESSGGL--GYFSDL-NF-----QAPVYVSTFDLSTIPSSLE- 51
DB 480 KADKALDREAKTTLGSLKHGVEFVDSNFYTNASLSPDKGVGATNG---VSHLEA 534
OY 52 -----NIPSENYFQSAIWSGFIKVKSDYTFATNSADNHV--WNVDD-----QEVIN 98
DB 535 GFSKVAAYFNPLNINLAITSY---VRODLEDKLAKGLSPOEAKHLYKDLSSNKKELG 590
OY 99 KASNSNRIELE--KGRLYQIKIOYRENPTEKGLFEKLYWTD--NKKKEVISDNLOPE 155
DB 591 KALNFKNAVAEAKNGONTDEVK--QAKODLEKSL-----KKRRELEKVAK--N 635
OY 156 LKQSSNRKSTKSTAGTVPDR-----DNDGIPDSLEVEGYVDVANKRTEFLSP 210
DB 636 LESKGNKNNMEKKSQANSQKDELFALINKKANRDAK--AYQNK--KGIKRELSQKLE 694
OY 211 HEKGLTKYSSPEKSTASDPISDEK-----YTGRI--DKNS--SPEARHPVAAPIV 262
DB 695 N--KDLADFSKSTDEKNGKN--KDFSKAETLAKLSYKVDGLGNPEIS-----741
OY 263 HYDMENITLSKNEDOSTONTSETRTYSKNTSTSTRTTSEYHGM/EVHASFIDIGSVSA 322
DB 742 --KVENLMAALNEFKNKNKDFSKVTOAK--SDLENSIKQVITLTKITDKVNLNQAVSY 797
OY 323 GFSNSNSTYVAIDHSLSLAGERTWAEETGTLNADTARLANANRYVNG--TAPIVNVLPTT 381
DB 798 AKATGDFSGV-----EOLALADLKNFSKEOLAQQAOKNED--NFGKNSALVO-----842
OY 382 SLVGLKQNOTATIKAKENOLSOILAPNNYPSKNIAPT--ALNAQ--DDE-----S 428
DB 843 SKYNKVNNGTLYV---GNGLSKAEATY--LTSKNFSQDKKRELAKIGNNNNNNGLENS 894
OY 429 SPTPTMAYNOFLEETKQLRLTDVGYGNATYFENGKRVYDGSWSEFVLPOIETT 488
DB 895 TEPV---YFQVAKYKAKIDRL--DQIASGLG-----DVGOAASFLLRHDKV- 937
OY 489 ARIINCKDLNVERIAVNPSPLETKPPM---TIKEALK--AFGPNPNCNLOYOG 544
DB 938 -----DLS-----KVLGSANHEPIYATITDLDGGFFLKRHKV-----971
OY 545 KITEFEPNPDQOSTQNTKN--QIAELNNT---NIITYLDKIKLNKAKNMLIRKRFHY 598
DB 972 DDLSSVGLSKEQTLQKIDNINQAVSAPKASHFDNLDQKIDKLSKTKNNV-----1023
OY 599 DNNNTANVAGADEVYKAREVYNSSTEG--LLNLT--DKQIDRKG1NGIYEIETDEGLAE 654
DB 1024 ---NLVYESAKKVPYSLSAKLDNATNASHRTNSVKNKGTINKAKGLQTO--KNSEWLK- 1078

```

OY 655 VINDRDMANISSLRQDGTETDF-----KRYND-KLPLYISNP-----NKYANV 698
Db 1079 LVNDKIVAHNVGSAPLSAYDKIGFNOKNMKDYSDSEFSTRLSNAVKDIKSGFVQFTNI 1138
OY 699 YAVTKENTLIINSENGDSTNGIKILIFSKRYE 733
Db 1139 ESMGSSYSIMKASVBEHGVKNTN-----TKGQO 1165

Search completed: October 4, 2002, 14:39:21
Job time: 271 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 14:35:25 ; Search time 34.15 Seconds

(without alignments) 3728.383 Million cell updates/sec

Title: US-09-848-909-21

Perfect score: 1 EVKQENRLNLSSESSQGL.....STNGIKLILPSKKEITGZ 736

Sequence:

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea:.*
2: SP bacteria:.*
3: SP fungi:.*
4: SP human:.*
5: SP invertebrate:.*
6: SP mammal:.*
7: SP plant:.*
8: SP organelle:.*
9: SP phage:.*
10: SP plant:.*
11: SP rodent:.*
12: SP virus:.*
13: SP vertebrate:.*
14: SP unclassified:.*
15: SP virus:.*
16: SP bacteriophage:.*
17: SP archaea:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3702	98.0	741	2	0937W2 bacillus an
2	3702	98.0	741	2	0937W3 bacillus an
3	890.5	23.6	876	2	032739 clostridium
4	890.5	23.6	876	2	006498 clostridium
5	888.5	23.5	876	2	09KH41 clostridium
6	884	23.4	875	2	046221 clostridium
7	826	21.9	721	2	086171 clostridium
8	200.5	5.3	2276	2	0937W6 staphylococ
9	199	5.3	4688	16	09PQ08 ureaplasma
10	188	5.0	2647	5	09U4X0 plasmidium
11	186.5	4.9	2771	5	026216 plasmidium
12	186	4.9	1227	16	097K41 clostridium
13	183.5	4.9	2529	16	025579 helicobacte
14	183	4.8	2747	5	09BJX9 plasmidium
15	183	4.8	4152	2	09ZHL3 haemophilus
16	182.5	4.8	1125	16	098PM9 mycoplasma

17	181	4.8	2269	5	026223 plasmidium
18	180.5	4.8	149	2	09RM77 clostridium
19	178.5	4.7	149	2	09RM78 clostridium
20	178.5	4.7	1072	16	09CF64 lactococcus
21	178.5	4.7	1387	5	09GZ76 plasmidium
22	178	4.7	752	16	098P19 mycoplasma
23	178	4.7	1302	2	049547 mycoplasma
24	178	4.7	2178	2	046149 clostridium
25	178	4.7	4919	2	09ZHL0 haemophilus
26	178	4.7	6713	16	09U4X0 staphylococ
27	178	4.7	6713	16	09U4X0 staphylococ
28	176.5	4.7	3254	2	09RM80 clostridium
29	176.5	4.7	3254	2	09RM80 clostridium
30	176	4.7	1193	2	045914 clostridium
31	174.5	4.6	3130	5	09BK46 plasmidium
32	174.5	4.6	3130	5	09BK46 plasmidium
33	173.5	4.6	821	3	008581 saccharomyc
34	172	4.6	1193	2	071107 clostridium
35	171.5	4.5	1939	5	025662 plasmidium
36	171.5	4.5	1939	5	025662 plasmidium
37	171.5	4.5	1365	2	049525 mycoplasma
38	171.5	4.5	2399	16	09ZKS9 helicobacte
39	171	4.5	1176	2	09F231 helicobacte
40	170.5	4.5	1197	2	045888 clostridium
41	170	4.5	1837	3	074424 schizosacch
42	170	4.5	2752	5	09BJX9 plasmidium
43	169.5	4.5	786	5	077357 plasmidium
44	169.5	4.5	144	2	049545 mycoplasma
45	169.5	4.5	2340	16	09ZD91 rickettsia

ALIGNMENTS

RESULT	ID	Q937W2	PRELIMINARY:	PRT:	741 AA.
AC	Q937W2	01-DEC-2001 (TREMUREL. 19, Created)			
BT	01-DEC-2001 (TREMUREL. 19, last sequence update)				
DT	01-DEC-2001 (TREMUREL. 19, last annotation update)				
DE	PGM PROTEIN (FRAGMENT).				
GN	Bacillus anthracis.				
OS	Bacillus anthracis.				
OC	Plasmid PKOL.				
CC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_Taxid=1392;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-FERRAR:				
RA	Adone R., Pasquali P., La Rosa G., Marianelli C., Muscillo M.,				
RA	Fasanello A., Francia M., Ciuchini F.;				
RT	*Sequence analysis of the genes encoding for the major virulence				
RT	factors of bacillus anthracis vaccine strain 'Carbapap' *;				
RU	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AJ413937; CAC93935.1; -				
KW	Plasmid.				
FT	NON TER				
FT	NON TER				
SQ	SEQUENCE	741 AA; 83153 MW; C7P95820E73065C0 CRC64;			

Query Match	98.0%; Score 3702; DB 2; Length 741;
Best Local Similarity	99.7%; Pred. No. 2.2e-176;
Matches	720; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 EVKQENRLNLSSESSQGLGYPSDLNFOAPVYVTSSTGDLSPSSLELNTFSENGYR 60
DB	20 EVKQENRLNLSSESSQGLGYPSDLNFOAPVYVTSSTGDLSPSSLELNTFSENGYR 79
OY	61 QSAIWSGFIKKKSDYETPATSADNHYVMVDDEYIKRASNSKRLKGLGYQIKYQ 120

```

Db      80 QSAIMSGEIKYKKSDEYTFATSDNHYTMVDDQEVYKASNSKIRLEKGRLYOIKIY 139
Oy      121 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLOLPELKOKSSNRKRSAGPYTPDDN 180
Db      140 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLOLPELKOKSSNRKRSAGPYTPDDN 199
Oy      181 DGIPLSEYEGYTVDYVKNRFTLSPWISNHEKGLTKYKSSPEKWSADPYSDPEKYT 240
Db      200 DGIPLSEYEGYTVDYVKNRFTLSPWISNHEKGLTKYKSSPEKWSADPYSDPEKYT 239
Oy      241 GRIDKNSVPEARHPLVAAPPIVHVDMENTILSKNEOSTONTDSEFTTSKTSRTH 300
Db      260 GRIDKNSVPEARHPLVAAPPIVHVDMENTILSKNEOSTONTDSEFTTSKTSRTH 319
Oy      301 SEVHGNAEYHASFDDIGGSVSGFNSNSSTVAIDHSLSLAGEPTAETMGLNTADTARL 360
Db      320 SEVHGNAEYHASFDDIGGSVSGFNSNSSTVAIDHSLSLAGEPTAETMGLNTADTARL 379
Oy      361 NANIRVNTGAPLYNVLPFTSLVYGNKQTLATIKAKENOLSQLAPNNYPSKMLAIA 420
Db      380 NANIRVNTGAPLYNVLPFTSLVYGNKQTLATIKAKENOLSQLAPNNYPSKMLAIA 439
Oy      421 LNAODFSSPTITMNTNOFLELEKTKOLRLDTPQYGNATYFNENGRVAVDTGSMSSEV 480
Db      440 LNAODFSSPTITMNTNOFLELEKTKOLRLDTPQYGNATYFNENGRVAVDTGSMSSEV 499
Oy      481 LPOIOETTRARIIFNGKDLNVERRIAANPSDPLETTKPDMTLKEALKIAFGFNPNCNL 540
Db      500 LPOIOETTRARIIFNGKDLNVERRIAANPSDPLETTKPDMTLKEALKIAFGFNPNCNL 559
Oy      541 QYQKDIETEFNFDQTSQNIKNOLAELNATNIYVLDKIKLNKAKNNILIRDKRFHHR 600
Db      560 QYQKDIETEFNFDQTSQNIKNOLAELNATNIYVLDKIKLNKAKNNILIRDKRFHHR 619
Oy      601 NNIAVGADESYYKKAHREYVINSSTEGLLNIDKIRKILSGIYEIEDTGLKEVYINDRY 660
Db      620 NNIAVGADESYYKKAHREYVINSSTEGLLNIDKIRKILSGIYEIEDTGLKEVYINDRY 679
Oy      661 DMNITSSLROGKTFIDFKKYNKDLPLXISNPNYKVNVAATKNTIINSENGDTSTNG 720
Db      680 DMNITSSLROGKTFIDFKKYNKDLPLXISNPNYKVNVAATKNTIINSENGDTSTNG 739
Oy      721 IK 722
Db      740 IK 741

```

```

SQ      SEQUENCE 743 AA: 83363 MW: 292757AD2D5D5E5A6 CMC64:
Query Match 98.0%; Score 3702; Db 2; length 743;
Best Local Similarity 99.7%; Pred. No. 2,2e-178;
Matches 720; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Oy      1 EVKQENRLNLSSESSSSGGLGYFFSDLNFOAPMVYTSSTGDLSPSELEINPSENGYF 60
Db      22 EVKQENRLNLSSESSSSGGLGYFFSDLNFOAPMVYTSSTGDLSPSELEINPSENGYF 81
Oy      61 QSAIMSGEIKYKKSDEYTFATSDNHYTMVDDQEVYKASNSKIRLEKGRLYOIKIY 120
Db      82 QSAIMSGEIKYKKSDEYTFATSDNHYTMVDDQEVYKASNSKIRLEKGRLYOIKIY 141
Oy      121 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLOLPELKOKSSNRKRSAGPYTPDDN 180
Db      142 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLOLPELKOKSSNRKRSAGPYTPDDN 201
Oy      181 DGIPLSEYEGYTVDYVKNRFTLSPWISNHEKGLTKYKSSPEKWSADPYSDPEKYT 240
Db      202 DGIPLSEYEGYTVDYVKNRFTLSPWISNHEKGLTKYKSSPEKWSADPYSDPEKYT 261
Oy      241 GRIDKNSVPEARHPLVAAPPIVHVDMENTILSKNEOSTONTDSEFTTSKTSRTH 300
Db      262 GRIDKNSVPEARHPLVAAPPIVHVDMENTILSKNEOSTONTDSEFTTSKTSRTH 321
Oy      301 SEVHGNAEYHASFDDIGGSVSGFNSNSSTVAIDHSLSLAGEPTAETMGLNTADTARL 360
Db      322 SEVHGNAEYHASFDDIGGSVSGFNSNSSTVAIDHSLSLAGEPTAETMGLNTADTARL 381
Oy      361 NANIRVNTGAPLYNVLPFTSLVYGNKQTLATIKAKENOLSQLAPNNYPSKMLAIA 420
Db      382 NANIRVNTGAPLYNVLPFTSLVYGNKQTLATIKAKENOLSQLAPNNYPSKMLAIA 441
Oy      421 LNAODFSSPTITMNTNOFLELEKTKOLRLDTPQYGNATYFNENGRVAVDTGSMSSEV 480
Db      442 LNAODFSSPTITMNTNOFLELEKTKOLRLDTPQYGNATYFNENGRVAVDTGSMSSEV 501
Oy      481 LPOIOETTRARIIFNGKDLNVERRIAANPSDPLETTKPDMTLKEALKIAFGFNPNCNL 540
Db      502 LPOIOETTRARIIFNGKDLNVERRIAANPSDPLETTKPDMTLKEALKIAFGFNPNCNL 561
Oy      541 QYQKDIETEFNFDQTSQNIKNOLAELNATNIYVLDKIKLNKAKNNILIRDKRFHHR 600
Db      562 QYQKDIETEFNFDQTSQNIKNOLAELNATNIYVLDKIKLNKAKNNILIRDKRFHHR 621
Oy      601 NNIAVGADESYYKKAHREYVINSSTEGLLNIDKIRKILSGIYEIEDTGLKEVYINDRY 660
Db      622 NNIAVGADESYYKKAHREYVINSSTEGLLNIDKIRKILSGIYEIEDTGLKEVYINDRY 681
Oy      661 DMNITSSLROGKTFIDFKKYNKDLPLXISNPNYKVNVAATKNTIINSENGDTSTNG 720
Db      682 DMNITSSLROGKTFIDFKKYNKDLPLXISNPNYKVNVAATKNTIINSENGDTSTNG 741
Oy      721 IK 722
Db      742 IK 743

```

```

RESULT 3
O32739 PRELIMINARY: PRT; 876 AA.
AC O32739;
DT 01-JAN-1998 (TREMblrel. 05, Created)
RT *Sequence analysis of the genes encoding for the major virulence
RT factors of bacillus anthracis vaccine strain 'Carbopap'.
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DE ADP-RIBOSYLTRANSFERASE.
GN CDTB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
NCBI_TaxId=1496;

```

```

RN      [1]
SEQUENCE FROM N.A.
RP      STRAIN-CD196;
RC      MEDLINE-97230316; PubMed-9119480;
RA      Perelle S., Gilbert M., Bourlioux P., Corthier G., Popoff M.R.;
RT      "Production of a complete binary toxin (actin-specific Adp-
RL      ribosyltransferase) by Clostridium difficile CD196."
DR      Infect. Immun. 65:1402-1407(1997).
DR      EMBL: L76081; AAB67305.1; -.
DR      HSSP: P13423; IACC.
DR      InterPro: IPR003896; BinaryToxB.
DR      PRINTS: PRO1391; BINARYTOXINB.
KM      Transferase.
SQ      SEQUENCE 876 AA; 98797 MW; 25E06E2D45CE2B3B CRC64;

Query Match      23.6%; Score 890.5; DB 2; Length 876;
Best Local Similarity 31.8%; Pred. No. 5,7e-37;
Matches 253; Conservative 139; Mismatches 278; Indels 125; Gaps 30;

QY      4 QENLLNESESSOGLGYFSDLNQAPMVYVSTTGDLSPSELENIPSENOYFOS 62
DB      40 KKKIYMEDILFNKGLMGITFSDHKKDLKMAPIKDKGNLKEEKVKDKLDDKSDVAS 99
QY      63 AMSGFIKKKDETFATSDNHNVTMVDQEVINKASNSKIFLEKGRLYQIKYOR 122
DB      100 IRWGRILPSKDEYTLSTDRD-VLMQVNESTI--SNTLKVMKKGKRYKRYLEI 155
QY      123 EN--PTKGLDPRKYWTDSONKKEVISDNQLPELKOKSSRRKSTAGPTV 176
DB      156 KNLGSDNLSSPNLW--ELDGMKKIIPENLELDY---SNEKD---DPIFNNEF 205
QY      177 -----DRNDGIPDLSLEVGYVDVKNKRTFLSPWISNHEKKGLTYKSS 222
DB      206 FDPRLMDWEDDLPTDNDNIPDSYERNQYTI---DLIAVAKMDSFAE--QGGKKYYSN 260
QY      223 PEKMSIADPYDEFEKYVGRIDKNVSPARHPVLAAPYVHVMENILSKNDQST 282
DB      261 YLESTAGDPYDYDKASGSPDKAIKEARDPLVAAPYVGVMKKLITSTNEHAST 317
QY      283 DSERTTSKSTSTRTSTSEVHGNAEVAHASFDDIGSVSAGFSNSNSTVAIDHS 337
DB      318 -DQKITYSRATNSKTESNTA-----GVSANVGQNGFTANVTYNTSHFTON 363
QY      338 ---LSLAGEETAETMGANTADARLANIRVYNTGAPTYNLTPTTSLVGGKQTLATIK 394
DB      364 STAVQDSNGESMWTGISTNGESYINANVRYNTGAPTYKTPPTLVV--DQDLITSTI 422
QY      395 KAKENQISQILAPNNYTPSKNLAPILANADDESSFTYNTNNOLELEKTKQLRLTDQ 454
DB      423 KAOENQIGNNLSGOTYFKKGLSPALNTMOFSSRLIPTNDQKLDAKQIKLETTO 482
QY      455 VYGNIAIYFNGRVRVDTGSNNSEVLPOIOTETARILFNGKDLNVERIAAVNDSPL 514
DB      483 VSGNFGTKN--SSGQI--VTEGNSMSDVISQIDISASIIIDTER--BSYERRVTAKNLQDE 539
QY      515 ETTPEDTTLKALKAIFGNEPENGNILOYOKDITE--FDFNEQOOTSONIKNOIAELNAT 572
DB      540 DKT--PELTIGELIEKAFGATRKKGGLLYFNDIPIDESCVELIFEDNTANKIDSKLISDK 598
QY      573 NIYVLDKIKLANKNILIRDKRFHY---DRNNIAVGADESVYKFAHVEVNSSTGL 627
DB      599 KIYVW---KLEGNMILIKPTPTFNFDDYNNP--STMSVNTTNOGGJOSANKLNG 652
QY      628 -----LNIQDKRLKILSGY-----IYEI-----EDDEGI-----KEYI 656
DB      653 EFKIKIPNSMLAPKYRVYFSGYSKDLPTNSLIYKIKAKKEKTDVLVEQGYTFSEYE 712
QY      657 NDRDMNLIS--SLRODKTFIDFKKTD--KLPLIXNSPKYK-----NYVAAYKENT 706
DB      713 TTEKDSNIEFTILSGFTYIDNLSTELNSPELIDEPYKIPFDQIDMAKHTFADL 772
QY      707 IINSENGDTSTNGI 721

```

```

DB      773 NFNP--TGNTYINCA 786

RESULT 4
ID 006498 PRELIMINARY; PRT; 879 AA.
AC 006498;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SB COMPONENT.
GN SBS.
OS Clostridium spiroforme.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
OX NCBL_TaxID-29348;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CD246;
RA Gilbert M., Perelle S., Daube G., Popoff M.R.;
RT "Clostridium spiroforme toxin genes are related to C. perfringens iota
RL toxin genes but have a different genomic localization."
RT Syst. Appl. Microbiol. 20:357-347(1997).
DR EMBL: X97969; CAAB612.1; -.
DR HSSP: P13423; IACC.
DR InterPro: IPR003896; BinaryToxB.
DR PRINTS: PRO1391; BINARYTOXINB.
SQ SEQUENCE 879 AA; 98739 MW; 40685ACB8E05BA01 CRC64;

Query Match      23.6%; Score 890.5; DB 2; Length 879;
Best Local Similarity 28.9%; Pred. No. 5,7e-37;
Matches 261; Conservative 147; Mismatches 257; Indels 237; Gaps 34;

QY      4 QENLLNESESSOGLGYFSDLNQAPMVYVSTTGDLSPSELENIPSENOYFOS 62
DB      36 QEVTTNEKTVSNTKMGYGFADHKKDLKMAVKKNGELKEEKVKDKLDEKKNITS 95
QY      63 AMSGFIKKKDETFATSDNHNVTMVDQEVINKASNSKIFLEKGRLYQIKYOR 122
DB      96 IRWGRILPSKDEYTLSTDRD-VLMQVNE--GEIANTLKVMKKGQYSRIEIO 150
QY      123 ENPTKGLDPE-----KIYWTDSONKKEVISDNQLPELKOKSSN----- 162
DB      151 ----DKDIGYDLSRKLY--ELNDGKTLIPKKNLFLNDYSKIDENPFIPKDNFEDK 205
QY      163 --SKKRSNSA--GPTVPDRNDGIPDLSLEVGYVDVKNKRTFLSPWISNHEKKGLTY 219
DB      206 LKSRGASRLASGMDDEDLPTDNDNIPDAIEKNGITL-----KQSLAVAK--EDSPQGGYK 260
QY      220 KSPKMSIADPYDEFEKYVGRIDKNVSPARHPVLAAPYVHVMENILSKNDOST 279
DB      261 LSTYLESNTAGDPYDYOKASGSPDKAIKEARDPLVAAPYVGVMKKLITSTNEHAST 320
QY      280 QNTDSERTTSKSTSTRT--HTSEVHGNAEVAHASFDDIGSVSAGFSNSNSTVAIDHS 337
DB      321 ----DQKITYSRATNSKTESNTA--GVAIINAYONGF---TGSITTYNVTSHTESTAVONS 373
QY      338 LSLAGEETAETMGANTADARLANIRVYNTGAPTYNLTPTTSLVGGKQTLATIK 397
DB      374 ---GGE--SNMTLSLSTNGESYINANVRYNTGAPTYKTPPTLVV--DQDLITSTIKQ 428
QY      395 ENQISQILAPNNYTPSKNLAPILANADDESSFTYNTNNOLELEKTKQLRLTDQYVG 457
DB      429 DNQIGNNLSFNETYFKKGLSPALNTMOFSSRLIPTNDQKLDAKQIKLETTOYVG 488
QY      458 NIAYVFNENGRVRVOT--GSNNSEVLPOIOTETARILFN--GKDLNVERIAAVNDSPL 515
DB      489 N---YGIKNSQOQIITBEGNSMSDYISQIDISASIIIDTGSQ--VEERVYATDSSNP 543
QY      516 TTKEDMLKALKAIFGNEPENGNILOYOKDITE--FDFNEQOOTSONIKNOIAELNAN 573
DB      544 KT--PVLTIGERATKAFGATRKNGELIFNGMPTIDESCVELIFEDNTANKIDSKLISDK 602

```

```

OY 574 IYTVLDKIKLNKNNILIRD----- 593
DB 603 IYVY-----QLEGRKILITSTYFNNFGYNNPPSSKSNVNDNNOGLOANAKLSGETK 658
OY 594 -----KRRHY----- 598
DB 659 IYVPSKLNPKRRYVSGYLNKSNSTNPTIVNINAKBQTYVLYSENDYKRSFEFEIG 718
OY 599 -DRNN-----IAGADESVYKAREV-----I 620
DB 719 RDSNNEITLSSSTIFLNDLSTELNSTPEILKEPDILVPSDDLI-DAKKKYADLSF 777
OY 621 NSSTEGILLN-----IDKDIRILSGYIVLE-DTEGLEVINDRYMLNLSLROD 671
DB 778 NOSTANYIYDGLYPEPTQNKREVLIDYIOKRYEATLYESGFDIGTOKDELRYTGDSDNQ 837
OY 672 GKT-FTDEKRYNDKPLIYISNPNY-----KVNVAVTRENTIINPENGDTSTGIRKILI 726
DB 838 PKTNYVNFHSY-----FTSGENVMPYKRIYALIPEN-----KELLY 875
OY 727 FS 728
DB 876 LS 877

RESULT 5
OY 598HAI PRELIMINARY; PRT: 876 AA.
AC 098HAI;
DB 01-OCT-2000 (TREMblrel. 15, Created)
DB 01-OCT-2000 (TREMblrel. 15, last sequence update)
DB 01-DEC-2001 (TREMblrel. 19, last annotation update)
DB CDB.
GN CDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CCUG 20309;
RA Chang S.Y., Song K.P.;
RT "ADP-ribosylating Binary Toxin Genes of Clostridium difficile strain
CCUG 20309."
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR HSP: A8271719; A881761.1;
DR HSP: P13423; JACC.
DR InterPro: IPR003896; BinaryTox.
DR PRINTS: PR01391; BINARYTOXINB.
SQ SEQUENCE 876 AA; 98793 MW; 366D62F352E745A5 CRC64;

Query Match 23.5%; Score 888.5; DB 2; Length 876;
Best Local Similarity 31.7%; Pred. No. 7.2e-37;
Matches 232; Conservative 140; Mismatches 278; Indels 125; Gaps 30;

OY 4 QENRLNSESSESSGGLIGYFSDLNFOAPVYVTSSTTGDLISPSSELEN-IPSENOYFOS 62
DB 40 KKEIYNEDILPPNGLMGYFTDEFKDLKMAPIKIDGNLKEEKEDKLDKSDVKS 99
OY 63 AMSGFKVKSDEYFATSADNHTVMVDDOEVAKNSRKILKGLVQIKIOTYR 122
DB 100 IRTWGRILPSKDGCVLTSTDRD-VLMQVNTSTI---SNTLKVMMKKGKEYVRIEOD 155
OY 123 EN--PEKGLDFKLYWTDSONKKEVYISDNLOJPELKOKSSNRKRKSTAGATP--- 176
DB 156 KNIGSIDNLSNPYVW-ELDKMKKILIPENLFLRDY---SNJEKD-----DPIPIPNNF 205
OY 177 -----DRNDOGIPDSLEVGATVYVKKRPTFLSPMISNIIHKGLTKRYS 222
DB 206 FDPKLMSDWEDEDIDTDNDNIPDSYERNGYTI---KDLINKWEDSFAL-GQYKRYYSN 260

```

```

OY 223 PEKMSADPYSDPEKTYGIRKNVSPKRRPIVAAPIVAVDHNILISKNEOSTONT 282
DB 261 YLESNADPTDYDEKASGSEFDKAKTEADPLAAVPIVGVGKELISTENHST--- 317
OY 283 DSEFTIKNTSTSRHTSEVGHNAEVAHFEDIGSGVSAGFSNNSSTVAIDHS----- 337
DB 318 -DQKTVSRATNTKTESNTA-----GVSVNGVONGFTANTYNSHTDN 363
OY 338 --LSIAGERVATMGLNTADTARLANIRVNTGTAPITNV-PTSLVLKNOTLATI 394
DB 364 STAVDPSNGESMNGLSINKGESAVIANVRYNGTAPMYKVPPTNVL-DGPTLSTI 422
OY 395 KAKENOLQILAPNNYPSKNIAPALNAODFSSTPTMNYNOJLEKTRQLRLTDQ 454
DB 423 KAOENQIGNNLSPGDYTPKKGSLPALNTMDGFSRLPIVNDLQKRLDAGQIKLETQ 482
OY 455 VYGNATYVNPENGVRVYDGSNMSEVLPJOIETARIFPNCKD-LVERRIAVNSDPL 514
DB 483 VSGNFGTKN-SSGQI-VTEGNSMSDYISQDISASIIIDTEN-FSEYERRVAKMLDPE 539
OY 515 ETTKPMTEKALKIAPGPNENSLAQOKDITE--EDPNEDQISONIKKQALFENAT 572
DB 540 DKT-PELTIGALIKAPGATKKGILFNDIPIDECVELLFDNTHANKIKSLTISDK 598
OY 573 NIYTVLDKIKLNKNNILIRDKRRHY---DRNNIAGADESVYKAREVINSSTEGI-- 627
DB 599 KIYVY-----KLERGMNLIKTYFTYFNDYNNYP--STWSNVNFKDGLGSKANKLNG 652
OY 628 -----LLNIDKIRKILSGY-----IVETI--EDTETL-----KEVI 656
DB 653 ETKIKIPSEIKPKRYVSGSKNDPLNSIIVKIKAKKERTVYLVPOGYTKSEFE 712
OY 657 NDRYDMLNIS-SLRDOKTFIDPKRYND--KLPYISNPNYK-----VNYVAATEMT 706
DB 713 TTEKDSNIEITLIGSGTYLNLSTELNSTPEILDEPEVKIPJOEILDAHKIYPADL 772
OY 707 IINPENGDTSTNGI 721
DB 773 NFNPVS-TGNNTYINCM 786

RESULT 6
OY 046221 PRELIMINARY; PRT: 875 AA.
AC 046221;
DB 01-NOV-1996 (TREMblrel. 01, Created)
DB 01-NOV-1996 (TREMblrel. 01, last sequence update)
DB 01-DEC-2001 (TREMblrel. 19, last annotation update)
DB IOTA TOXIN COMPONENT IB PRECURSOR.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RA MEDLINE-94041637; PubMed-8225592;
RA Perelle S., Gilbert M., Boquet P., Popoff M.R.;
RT "Characterization of Clostridium perfringens iota toxin genes and
expression in Escherichia coli."
RL Infect. Immun. 61:5147-5156(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RA Popoff M.R.;
RT Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DB EMBL: X73562; CAA51960.1;
DB HSP: P13423; JACC.
DB InterPro: IPR003896; BinaryTox.
DB PRINTS: PR01391; BINARYTOXINB.
DB SIGNAL.
DB 34 38 POTENTIAL.
DB 212 875 IOTA TOXIN COMPONENT IB.

```


RESULT 8
ID 09376 PRELIMINARY; PRT: 2276 AA.
AC 09376;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BIOFILM-ASSOCIATED SURFACE PROTEIN.
GN BAP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V329;
RX MEDLINE=21189316; PubMed=11292810;
RA Cucarella C., Solano C., Valle J., Amorena B., Lasa I.,
RA Penades J.R.;
RT "Bap", a Staphylococcus aureus Surface Protein Involved in Biofilm
RT Formation.
RT J. Bacteriol. 183:2888-2896(2001).
DR EMBL: AF288402; AAK38834.1;
SQ SEQUENCE 2276 AA; 238792 MW; 8623C60472CB0C0F CRC64;

Query Match 5.3%; Score 200.5; DB 2; Length 2276;
Best Local Similarity 19.2%; Pred. No. 0.094; Mismatches 300; Indels 267; Gaps 39;
Matches 166; Conservative 131;

47 SSELINPSENQFOSAINSGFIKXKSDERTFATSDNHVYMWDOEVINKASNSKI 106
43 AAEIDITTEDEVKQDGEALDIKNIKESKDVTEDDNNAY-QNSAQVYKSENSNDT 101
107 RLEGRGLYQIKYORRNPTEKGLDFKLYWDSONKE-----VISDMLQPELK 157
102 AVESTN-DSVKDEKTESKNSAO-----DDNKEKSNQSESTRTSSOSSEVPQK 154
158 OKSNSKRRKRSAGPYPP---RONDGIPDSLEVEGYVYVYKRFPLSPATINHEK 214
155-KVNETSETAIDEDASTKQNNKNDNRADDNKKEDS-NQSESTRTSSOSSEVPQK 212
215 GLTYKSSPEKNSA-----SDPYSPFEKVTGRIDKNSPEARH 253
213-----KQDPKSSNRIKPEPKQOEYAKKEKAITETLADNKKELKNNKDKNESELE- 266
254 PLVAAPYIVHVMENIILSKNEDQ-----SPONTDETRITSKNTSTSRTHSEVHGA 307
267-----SNLSSENKKDVTESPLNSQLSDSEKTKIMEANADYDKATDEEINT 313
308 EV-HASFPIGGSVAGFSNSNSTVAIDHSLSLAGRTWATMGUADTARLANIRY 366
314 ELRLASLIEM-----ANNKKRTETL-----ATQPTMFRAATPTALRAVNODEL 360
367 VMTGAPYIVNPTSLVLGKNGTLATIKAKENQ-----SOL-----AP 407
361-----QKSLGYTDNTYFASMLPDPKLSDDLALNSNIIPDLSYSGANGS 407
408 NNYPSKNIAP-----ALNAODDESSPTI--TMVNYNOFLTEKTOLRL--DTPQVYG 457
408 NRYKIDKIDPITIAEHYTKISANPSGKKPVEFAKKNKNGNLDTQVEMFIRANDGLFG 467
458 N-----IATVYFENGGRVYDTSNMSEVLQIOETARLIFNGDL-----NIYE 502
468 GAEILSGYAKKNGKIELD-----DIVGNISNAGNLNNKLNHGVFVADSE 514
503 RRIAAVNPDPLETKRDPMTL-----KRAKLIANG--FNEPKNL----- 540
515 NKIVRSESSGFLKADDDLVLENNYSTENNAFASSGSAGATNENVEPFGILIDQ 574
541-----QOGKDITEEDFNFDDQ----- 557

575 IMKGFISYKTKANQWAVYQIDKDLPIYEGVELHGYDYKGLNGCDKRNKRYADL 634
558-----TSQNIKNQGLAEANNTIYVLDKI--KINAKN-NILTRKRRHYDRNIA 604
635 TIDEVNGNTISDML-NKLIERNMNLPEYGVAVYAKLNKSNVNIITKDAKDSGN--- 690
605 VQADESVYKREKREVIN-----STREGLLIMDKDIRKLSGYLVEIEDPBGAKVIN 657
691-----LIRETTCKEDPTFPAGYLDSGALIN-----NTLGSYLAQDYD--KQGL 736
658 DRID-MLNSSLRODGTFFDKRYNDKPLPYISNPNT-KVNVYAVKENTIIIPS---- 711
737 DREROLSLDAENE-DTGDGRKNGDEVVNYTSPVGKPOAAYITTEDVAVSGVPLK 795
712 ENGDTSTNGIKRILIFSRRGYEIG 735
796 EGAAOT-----AKVINAGCTVVG 814

RESULT 9
ID 09P08 PRELIMINARY; PRT: 4688 AA.
AC 09P08;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN U0482.
GN U0482.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROVAR 3.
RX MEDLINE=20500219; PubMed=11048724;
RA Class J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum".
RL Nature 407:757-762(2000).
DR EMBL: AB002145; AAF30894.1;
DR InterPro: IPR001152; Thymosin_B4.
DR Pfam: PF00240; Ubiquitin_1.
DR SMART: SM00152; THY; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFE1997E CRC64;

Query Match 5.3%; Score 199; DB 16; Length 4688;
Best Local Similarity 21.2%; Pred. No. 0.29;
Matches 173; Conservative 132; Mismatches 325; Indels 186; Gaps 40;

19 ILGYSPLNFOAPMVYTSFTGDISPSELENIPSENQFOSAINSGFIKXKSEY 78
3699 LVQVYLD-NIHONIDETRIKPDHNV-SKEIINPGVTMISKGNKSPDTVTANPEFK 3756
79 FATSADNHYTWVDDQEVINKASNSNRIKRGRLYQIKYORRNPTEKGLDPLKYWD 138
3757 IETQ-----DDNDVNLNIDATVAFKDEHNIRIKQIVRIKEN-----ND 3795
139 SQNKKREYISDNQQLPELKQSSN-----SRKKRSTSGAPTVPRD----- 179
3796 WLKQGI-----DNLN-PETVYKLENIELSKPLKHTNLSVSIENDKENISLITETGPNVLYK 3851
180-----NDGIPDSLEVEGYTDVYKRNKRTPLSPWISNIHEKKGLTYKSSPEKWSASDPYSD 225
3852 IGQNDTIDNPOQTIWITLSGVNSK-YNGRQIVVYKQNNNVYIYES--LITLQKQND 3907
236 FEKVTGRIDKNSPEARHPLVAAPYIVHVMENIILSKN-EDOSTONTSEI-RTISKRN 293

```

Db 3908 YQLLSNLNSN-----REYREKEIENHNSNTNNEDEKLNGVSNFTIQTQNT 3957
QY 294 STSRFTTS-EVHGNAEYHASF-----PDIGSASAGFS-----NSMS---STVAID 335
Db 3958 TVQMDSSATVIVGRGVNENFKIKSEDKILENNQOVAFPAKETIYDINWLOTRPLK 4017
QY 336 HSLSLAGETAAETMGNT---ADTAELMANIKRYVTGTATYIVLPTTSLVLCNQTLA 392
Db 4018 DVTSDKEGTMAHDLNSVNFKEETTYKLVIOFVAKPKKANNINSENNVILDTNISI 4077
QY 393 -----TIKANEOLSOILAPNNVYPSKNLAPALN-ADODESSPTIMNY--NOFLELE 443
Db 4078 NSVFEETTVGDHKLINTYSSNNVNTNQTINFTLSGVKMSVGRKTKLSYKSNQTSESI 4137
QY 444 KTKQLRLTDQYGNATYVNFENGK-----VAVDGSNMSEVLPOIQTETTAIIENGKD 497
Db 4138 HTEVEVLESNTQYINILNNLKRNTYTLIDVKLIDNNVNSDFKEENLTNSPTTSTSA 4197
QY 498 LNLVERRIAANVPSDPLETT-----KPDWTKEA-----LKIAFGNEPNCNLQ 541
Db 4198 INVNLTEELISNRASTNLKSTIIRKINLNDPQNVLRDQKQTYVGNKQAMGFIVSGNIK 4257
QY 542 YQCKDTEFDFENFDQOTSQNIK-NQLAELNATNIVYLDKIKLAKANNILIRDKRFHYD- 599
Db 4258 YLAATLVLDNEN-DKVNIVNISFNKRPISIAEN-----IGIKSNNTI-----YNDNS 4304
QY 600 -----RNINIVGA-----DSVYKFAHREVINSGSTEGILLINDKDIRKILSGYIIEDETE 650
Db 4305 IPRLEINNDIIVNGPIKKEIYVKNANKK--NNTIDVLDGLQINPRILAHNLK-PLAFKSTN 4361
QY 651 GKEVYINP--RYDMLNITSL-RQDKETPIDP-----KKY-----681
Db 4362 -----NOIEIVINSSIVNNDKSTISIRPTLANKKALYSLVQYIYVNNNSNTIVE 4415
QY 682 NDKLPLXISNRYVAVYVAVTKEKNTIINPSENDTS 717
Db 4416 SMLCP-KLNNINITYOK--INSHITL--SKNGEMS 4445

RESULT 10
QY04XO PRELIMINARY; PRT; 2647 AA.
ID Q904XO;
AC Q904XO;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE ERYTHROCYTE BINDING PROTEIN BBL-1 (FRAGMENT).
GN BBL-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20078864; PubMed=10613703;
RA Peterson D.S., Wellens T.E.;
RT "BBL-1, a putative erythrocyte binding protein of Plasmodium falciparum, maps within a favored linkage group in two genetic crosses."
RL MOL. Biochem. Parasitol. 105:105-113(2000).
RT EMBL: AF131999; AAB3018.1; -.
DR InterPro: IPR002828; Sure.
FT NON_TER 2647
SQ SEQUENCE 2647 AA; 304550 MW; AE98F8BD754E300 CRC64;

Query Match 5.0%; Score 188; DB 5; Length 2647;
Best Local Similarity 21.4%; Pred. No. 0.49;
Matches 183; Conservative 134; Mismatches 295; Indels 242; Gaps 46;

```

```

QY 43 LSPSELEENPESNPOYPOSATMGFLKYSKDEYTFATSDAHVIMWDDQEVY----- 97
Db 1133 SSGSSMSTESIRTDNDKQFT---SEDIAPINCHKEIGSSMDKCS---EDKSTIIDSE 1186
QY 98 ----KASNSKIRLE-----KGLYQIKIOYORNPRTKGLDRLKLTWD 138
Db 1187 NEENKSSHSDIKSDNKGSTIDYSLSESPKQDESY-----SPSSIDDLK----- 1234
QY 139 SONKKEVISS---DNLOPEIKOKSSNRKRSSTAGPYPRDND-----GLPDS 186
Db 1235 -PKKSSPVSFDVHDSFNISELSQASQNASDYQGERPSKNILRTDDVSEKREKISVSPN 1293
QY 187 LEVEGYVAVKNNKTEPLSPWISNIHERKGLTKYKSSPEK-----WSTASDPYDFEKY 239
Db 1294 VSY--TYDEDDKQKQISD-DSSIH-----HEIDPEKNLHV-SFSLSGLEBQDEKE 1342
QY 240 TGRIDKN--VSPERHPLVAAPYV-----HYDMENILSKNEDOSTONTDSE 285
Db 1343 KGEKESGLIPISPEKINDGKRENIYDPSYSEKVDNMRSDISNV--SEDESSILSR 1399
QY 286 TRTISKNTS--TSRTHTEVHGNAEYHASFIDGSVAGFS-----NSNSTVA 333
Db 1400 NTEGINSEFELASEEHTSVQVNNRDNHNRKQENLVSSSTQDESFRERREKENADSS-- 1456
QY 334 IDHSLSLAGERTAAETMGINTADTAR-----LNNIRYVNTGTA---PIYVNL-PTT 381
Db 1457 --HESELSSISEVGETIRRNQDAEASENDKGEIDILQSEGOIVTEKTKLEPTVINLQST 1514
QY 382 SLVLRKQNTLATIKAKEN--QLSOLAPNNVYPSKNLAPALNADODESSPTIMNYNQ 438
Db 1515 PLOGELHKRSELDKNDNGCVHTDANTFOYRBSSEVYITNKDKEMET--TKPESQ 1572
QY 439 FLELEKTYQLDLDQYVNT-ATYVFNENGRVAVTQSGMSEVLVQIOE-----T 487
Db 1573 TYKE-----IEIDSTRQDN--DQSN--SLIFLKNNEDEQVSPIS 1613
QY 488 TARIIPKNDLNLVERRIAANVPSDPLETTPKPMTKALKATVAGNE-PKNDLOYQCKD 546
Db 1614 TRVWMSG--SFVSRQEDIIIEKDKHIT-DDTITNPSENGKLGQEMPNDSI--KS 1665
QY 547 ITEFDENFDQOTSQNIKQLAELNATNIVYLDKI-KLAKANNIL-----IDKR 595
Db 1666 VT-----ITSPLRQVEQMIIEPIDGKGNENKNIIGQESSTEIR-KQ 1707
QY 596 FHYDRNNIAGV-----ADESVYKFAHREVINSGSTEGILLINDKDIRKILSGYI--- 645
Db 1708 MDGPISSVNIIEELHPVAGSKLEEKERSMDADQK--TITEDITV-----VVEDPBG 1758
QY 646 IEDTBGLKEVINDRQMLNLSLSLRQDKTFIDPKTKYNDKLPVYSNPNTKVAVYVTKEN 705
Db 1759 IGEHQIKVEHQAQSELNTYNTL-DGRIVVEYKERLIDENPGSLPNDRTITEHIELDEK 1816
QY 706 TIINPSE-----NGD 715
Db 1817 EIEHREHEDAHNGE 1830

RESULT 11
QY06216 PRELIMINARY; PRT; 2771 AA.
ID Q26216;
AC Q26216;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE RHOPTRY PROTEIN.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE OF 379-2771 FROM N.A.
RX STRAIN-YM;
RX MEDLINE=97077455; PubMed=8920022;
RA Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;

```

*Comparison of two members of a multigene family coding for high-molecular mass rhotry proteins of *Plasmodium yoelii*.*
 RT Mol. Biochem. Parasitol. 76:329-332(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YM;
 RX MEDLINE-20446215; PubMed-10989154;
 RA Green J.L., Holder A.A.;
 RT *Structure of the 88 gene encoding a high molecular mass rhotry protein of *Plasmodium yoelii*.*
 RT Mol. Biochem. Parasitol. 110:167-169(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YM;
 RA Holder A.A.;
 RT Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U36927; ABA41263.3;
 SO SEQUENCE 2771 AA; 325638 MW; C0CCB9AB6E7ACF36 CAC64;

Query Match 4.9%; Score 186.5; DB 5; Length 2771;
 Best Local Similarity 21.5%; Pred. No. 0.62; Mismatches 221; Gaps 43;
 Matches 180; Conservative 123; Indels 314;

```

OY 1 EVKQENRLNSESOGGLGTFSDLA-----FOAPVVTSTTGDI 45
DB 772 EYAEVQVNTYKTSIKLKHQNDINIKERKONQDQFEMHKTIPNENISK 831
OY 46 PSELENIPSENOFQSAI--WSGFIK--KSDERTYATSDNHYVMVDQEV--1 97
DB 832 TYNELIKMDE--FLSKVKNKYNDPKYKKEKSEHNFELTNKTKTSEELKYE 888
OY 98 NKASNSNKRLEKGLYQIKIOYRENPTEKGLDFKLYTDSQNKREYISDMQLPEIK 157
DB 889 NKFPDNRSLINTEK--SIEEYQNTMLKKVVD--YTKVCLNTNELTCHNKQTLK 943
OY 158 QKSSNSRK--KRTSAGPTVPDRNDGIPD-----SLEVEGYTVVKKRFTLSP 205
DB 944 DKLQNTKTKTKTSIDKITYTKFENILTDKKELETKFTGLSLNHNESNKKELLYEFD 1003
OY 206 WISNHEKKGLTKYKSSPEKSTASDPYSDEKVTGIDKNVSPAPRHPVAAVPIYVD 265
DB 1004 LKANLKKNNENLYKOFNEK--EKAVEIDIKKNDVIKYSN-----IEIT 1047
OY 266 MENIILSKNEDQSTQNTDSEFTTSKNTSTRTSTSEVHGNAEYHASPDIQGSAGPS 325
DB 1048 IYTSIYNIED-----TENE--IGKSIELNFTKVE-----KYAAVNT 1083
OY 326 NSNSSTVAL-DHSLSLAGEPTAETNGLTADTARLNIIRVYNTGAPLYNLTPTSLY 384
DB 1084 NLNEIKERKLDYDQDFGK-----EKNIKYPDEN--KIKNDIDTLNOK 1124
OY 385 LGRN-GLATFKA--KENOLSOI-----LAPY--NYVP--SKNLAPFA----- 420
DB 1125 IDSIELETKKNSHNHIDEIGDILKRVKPMFNEDEKIEKIKNIYEKIDKK 1184
OY 421 -LNAODDFSTPIYMNNOFLTEKTOLRLDPOYGNINAIYNEGRVAVYQNSSE 479
DB 1185 NIYEIDKLINISTENDKTSLEKKNINLSYKSGULFLOQIDKKKAHTIKAME 1244
OY 480 V----LPOOETARILFNKDLN-----YBRRLAVNSDP-----LET-----K 518
DB 1245 AYIDDLNKKKSOET--EKEMNIMDKIDIKKKKLNSHDQIKYIHTISKNEEK 1301
OY 519 PDMTLEAKIAFGNRP--NGNLYOGKDIYFPDNDQOTSQNTKNOIAELNATNY 575
DB 1302 ISDIKNSLKIQDSSESTINDIKELKAVLE-----SONNNDINOGLSKITE--NLY 1354
OY 576 TVLDKIKLAKANNILIDKRF--HYDRNNIAGVADSVYKAEHREVTNNSTEGLINIDK 633
DB 1385 NIL--KLKKIKKIIDVYKYEIDELKNN-----KRLMELNS-----1390
OY 634 DIRKILSYIVIEDTEGLKE-----VINDRY--DMLNLSLRODKTFTIDKRYND 683

```

```

DB 1391 --EKI-----IQLKENSLEEQOSKIKSTIDNYSSECIKNITML-----KTYIVENKNN- 1439
OY 684 KLELYISN-PYKRVN-----YAVKENTIIINPSENGDTSPN--GIKKILIFSKK 730
DB 1440 -INTYFKNAEYNOVNSLNNNIEWADTKSQYITLNKKNCTNNTQVNIKELEKREKK 1496

```

RESULT 12

```

O97K41 PRELIMINARY; PRT; 1227 AA.
AC 097K41;
DT 01-OCT-2001 (Trembl, 18, Created)
DT 01-OCT-2001 (Trembl, 18, last sequence update)
DT 01-OCT-2001 (Trembl, 18, last annotation update)
DE UNCHARACTERIZED PROTEIN, PROBABLY SURFACE-LOCATED.
CM CAC1080
OS Clostridium acetobutylicum
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_taxid=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE-21359325; PubMed-11466286;
RA Noelling J., Bretton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Nelson R., Lee H.M., Dubois J., Qiu D., Hilti J., Wolf Y.I.,
RA Ratunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007624; AN79054.1;
KW Complete proteome.
SQ SEQUENCE 1227 AA; 134031 MW; E6EBE3399AF6FB38 CAC64;

```

Query Match 4.9%; Score 186; DB 16; Length 1227;
 Best Local Similarity 19.0%; Pred. No. 0.23;
 Matches 167; Conservative 136; Mismatches 287; Indels 288; Gaps 44;

```

OY 37 SSTGSLIPSELENIPSENOFQSAI--SGFIKKSDEYTVYTSADNHYVMVDQEV 95
DB 14 SVTTSVILSKPAIKAVAADNSYKSSISNSNEINVKKE-----YQKE 58
OY 96 V-INKASNSNKRLEKGLYQIKIOYRENPTEK-----GDEKLYWDSQNKREVI 146
DB 59 VANNKNNDSNS-KVSSNENQVS--NKENSNKYS--SSSEIOSIKNNVNLQYQNNKSVL 114
OY 147 SSDNQLPELKQKSSNRK-----RSTNSAGPTVPDRNDGIPRSTLEVGTVDPYKRR 201
DB 115 AASVNDDEVKINSQVOTSDYIALGETKYKPTDILINKAIYVANSABG--TDLSEVEI 172
OY 202 F-----LSPWISN-----HEKG-LTKYK 220
DB 173 YDIYQTPALQAFNINDGVANVSDYTLGATYVNDANIDSVKTFYHKKYATVYK 232
OY 221 SSPKSTNA-----SPYSDEKVTGRIKNNVSPAPRHPVAA 258
DB 233 DVATKTSNALKNNINGOGGETTYTALVEGVTPTILD-----VKNNTVYKQ----- 280
OY 259 YPIVADNENIILSKNEDS--TQNTDSEFTTSK--NTSTSTRTSEVHGNAEYHASPDI 316
DB 281 -----NKGRDLITELSDSAASTARLNTALDN-----MDA 311
OY 317 GGSVAGSAGSNSSTVAIDHSL--SLAGERTVAETM-----GLNTADTARLNI-- 364
DB 312 GAVTLEDYQALGANNVPOHLVADVANSAMDQFQVSAIDGINTIMTYINNINSVGTE 371
OY 365 -RYVNTGAPF-----YVLEPTTSL--VLGNQTLATIKAKENOLSOILAPNNY----- 411
DB 372 DDYINSHAVDSNCGNIDYDILMANIIEKTKAKGDL--TIPEVANVYKEVKTLIDFYMHAA 430

```

```

QY 412 -----PSKNIAPIALNADQSSPTIMVNOFLEKTKOLRLDPOV----- 455
DB 431 AGOTTLQDVKRNDPAQVODDVA-----LSDMLTKRDKTKALQDQILSIINSK 484
QY 456 -----YCNITVFNFGVAVDGSNMSEVLPOIOETIARI-IFNGKDLNVERIAAVN 510
DB 485 INSGINIDDS-----KIQEAVDASKLEAVADIKKIKKADGRDLTIOELRSVKT 538
QY 511 SDPLETT-----KPDWTLKALKIAF-GNEPENGLOQOGKIDTEFDEPNDOOTSONIK- 563
DB 539 IDYNSTSVNGSGOSVDYITIGIDVTEI- NIEFVNERIKESGTLI- TIKMKIV 593
QY 564 -----NOLAEIN-ATNITYVLD-----KIKNAKKNILIDKRFHD 599
DB 594 IEPVQVLESEYVRYTVGVVVOYKTLGINNVNOMNIYINMELKKKQDYKIDDIOTFVD 653
QY 600 R--NNT-----ANGADESY-----YKAEHREYINSTEGLILIDKRIKLSGYIE 645
DB 654 NTINIDIVINKIGADVALSDYRIGIDYODILDYVNAADKIQNRYADVDILIEVKA 713
QY 646 IEDTEGLK-----EYINDRDMNLNSSLROGKTFIDFKKYDKPLV---ISNPNYK- 695
DB 714 ISSYEALMRINIGEAVIDFEKALGLDI-----NGILLIATITDLQNNKYT 760
QY 696 -----VNVY-AATKENTINPSENGDTSTNGI 721
DB 761 ADEVIAVQAQIEYRALMQIN-TAKATTADYNTLGI 796

RESULT 13
ID 025579 PRELIMINARY; PRT; 2529 AA.
AC 025579:
DB 01-JAN-1998 (TREMBLrel. 05, Created)
DB 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DB 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DB TOXIN-LIKE OUTER MEMBRANE PROTEIN.
DB HP0922.
DB Helicobacter pylori (Campylobacter pylori).
DB Helicobacter pylori; epsilon subunit; Helicobacter group;
DB Helicobacter;
DB Helicobacter;
DB NCBI_Taxid=210;
DB [1]
DB SEQUENCE FROM N.A.
DB STRAIN=26695 / ATCC 700392;
DB MDLINE=91394467; PubMed=9252185;
DB Tomb J.-F., White O., Kellavage A.R., Clayton R.A., Sutton G.G.,
DB Fleischmann R.D., Ketchum K.A., Klein H.P., Gill S., Dougherty B.A.,
DB Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
DB Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
DB McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
DB Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
DB Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Watney L., Wallin E.,
DB Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
DB Venter J.C.;
DB "The complete genome sequence of the gastric pathogen Helicobacter
DB pylori."
DB RT
DB RT Nature 388:539-547(1997).
DB RT EMBL: AE000602; AD07969.1;
DB RT TIGR: HP0922;
DB RT InterPro: IPR004311; VACA.
DB RT Pfam: PF03077; VACA.3;
DB RT Hypothetical protein: Complete proteome.
DB SEQUENCE 2529 AA; 274561 MW; 440828E8644472EC CRC64;
QY Query Match 4.9%; Score 183.5; DB 16; Length 2529;
QY Best Local Similarity 21.0%; Pred. No. 0.77;
QY Matches 156; Conservative 94; Mismatches 261; Indels 241; Gaps 36;
QY 123 ENPTEKIDPKXTDSONK-----KEVISQNLQPELKKSNSRRKRSAGAPVDPDR 178
DB 729 QNPRESVMSGYVNL--QNKTYNKGITYIID-----PILSGSGSGSNLSTYTNLF--- 779

```

```

QY 179 DNDGIPDLEVEGYVDVKNKRTFLSD---NI-----SNIEKKGLTKK 220
DB 780 ---GGSFVNIQNGTLIGLNNTNESVNGNLITGOGGFGYITGFSNANTY---LTNNF 832
QY 221 SSPEKMT-----ASD-----PYSDPEKRYGRIDKNVSEARHPLVAPV 262
DB 833 KQEGVNSDDGANTTFKASNTLMOGLYNDAEIVYTKMIOFGA---QHSYATFDALN 889
QY 263 HYDMELLISKNEDQSTONTDEFTISKNTSTFRFTSEVHAGNAEVAHSPDIDGSYSA 322
DB 890 NISVMSSES-----DMTWKFFSEKSNISES-----S 921
QY 323 GFNSNSSYVA---IDHSLSLAGEPTAETMGNTADTARLANALRYVNGAPLYNVL- 378
DB 922 GFTNPGGSSVISAANTSLSFINS-----LNGAVYVNLQANS-LEN-NQAVFVNLV 973
QY 379 -----PTSLVKGNOTLATIKAKENQSLQILAPNNVPSKN-APIAL--NADQDS 428
DB 974 SRTGSNPNATVQLGNTNFTLS-----QSLNFGNDITLNNANITLIGKKSQAIF 1025
QY 429 STPTVNNYNOFLEKTKOLRLDPOVGNATYFENGRAVAVDTSSNMSEVLPOIOETT 488
DB 1026 NS-LTLDNNSNLSDNOSVLANNTSAPFNQASLNTYNS-----QATF 1068
QY 489 ARIENKDLNL-VERIAVNPSPLETTKPDWTLKALKIAFENPENGLOQOGDI 547
DB 1069 NSLFPNGCTLSNASSKINSNPSFNMT---INDDSVLSASVTSLSNANINFGASQ 1125
QY 548 TEF-----DEFNDOOTSONIKNOLA-----E 568
DB 1126 ADPGNTIIDTASFNDSASLNNLITANGALNNGYTPSLTKALMSVSGQVILNCGD 1185
QY 569 INATNITYVLDKIKNAKKNILIDKRF-----HYDRNNIAY 605
DB 1186 INLSDI-NITDNTSAYTNLNAQGITGIGANGYEKILFYCKIKQNTATSPNNNDIT 1244
QY 606 GA-----DSVYKAAHR-----EYIN--STSEGLINIDKDI--RKILSGYIE 645
DB 1245 WFLNPLNSQIIQIESIKNGDLTEVANNPNSASNTINIAPELVYQASQNPQTSVD 1304
QY 646 IEDTEGLKRYNRYMNLNSSL-----RODKTFIDKRYNKLPLY---ISNPNY 694
DB 1305 YSDNQA-----GYTLTSNKGLETPPGSGTPOAGTSPENQPLSSLNINYKKSSENL 1359
QY 695 KVNVAAYVTKENTYIN---PSENGDTSTNGIKIL 725
DB 1360 KTLIGITSONSATLKEMIESNODNITN-INEVL 1392

RESULT 14
ID 09BJX9 PRELIMINARY; PRT; 2747 AA.
AC 09BJX9:
DB 01-JUN-2001 (TREMBLrel. 17, Created)
DB 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DB 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DB 235 KDA RHOPTRY PROTEIN (FRAGMENT).
DB IIA.2.
DB Plasmodium yoelii yoelii.
DB Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
DB NCBI_Taxid=73239;
DB [1]
DB SEQUENCE FROM N.A.
DB STRAIN=YM;
DB MEDLINE=21273138; PubMed=11378199;
DB Khan S.M., Jara W., Bayele H., Preiser P.R.;
DB Distribution and characterisation of the 235 kDa rhoptry multigene
DB family within the genomes of virulent and avirulent lines of
DB Plasmodium yoelii.
DB Mol. Biochem. Parasitol. 114:197-208(2001).
DB EMBL: AF223443; AA15626.1;
DB FT
DB NON_TER 1

```

FT NON_TER 2747 2747
SQ SEQUENCE 2747 AA; 321237 MW; 430508AC7B8624BD CRC64;

Query Match 4.8%; Score 183; DB 5; Length 2747;
Best Local Similarity 19.1%; Pred. No. 0.91;
Matches 162; Conservative 132; Mismatches 332; Indels 220; Gaps 36;

QY 1 EVKQENL-----LNESSSGILGTYFSDLANFOAPVYVTSSTGDLSPSELENIPSE 56
DB 960 QTKLENTFTESLNHNANNEELI-KTESDLAKANGINEENMLYNOFTEKTEPNDIEK 1018
QY 57 NOYQASLNSGPIKAKSDERTPATSDNHYTAWD-----QEVINKASNSKTRLEK 110
DB 1019 NIHNEELIKIEKTHAS-IYNISETEREIGINISLNTKVEEYKKNVYLNKREK-ER 1076
QY 111 GRLYQI-----KIQY-----QRENTEKGLDFLYWDSQNKKEVISS 148
DB 1077 LKHDFSDGKSGNIRKYDKIRKINDIYAVSOQIDQIHGIDL-----DIORSSSEYVS 1130
QY 149 DNIQLPELAKSSNSRKRSTAGPTVPDRNDGIPDSLEVEGYTVYKNNKTFELSPMS 208
DB 1131 EMEXKXKLEKVSNT-----EISND-----NVEG-----IKKQOII---VT 1164
QY 209 NIHEKKG-----LYKKSPEKMSASDPYSDPEKYTG-----RIDKNVSPPEAR 252
DB 1165 KIDKKKIYEEIKKLSEIKIEKDNLSLEKYDINISYGOINIGLPLEQIDEE-KKAE 1223
QY 253 HPLVAAPPIYHDMENILSKNEDOSTQWDSERTISKTSTSRHTSEVHNAEVH-- 310
DB 1224 NTIKSMAYID-DLNIPIKKSQIEETEMDIKMDINKMEMLKISHDDKKCHDKSNHME 1282
QY 311 --ASFPGIGSASGFSNSNSVAIDHSLSLAGERWAMTGLTAPATRLANRIYVN 368
DB 1283 NISDIYKSSKLIQOFSE-ESDINDIKMLQ-----KIVSSQNHSDINQCL 1329
QY 369 TGTPIYNYLPTSLYLOKNOTLATIKKENOLSLAPNNYPSK-----N 415
DB 1330 MEVANIYIL-----KLNKIKIIDKVEYTSIEIKKKKNINDELNN 1371
QY 416 LAPALMAODSSPTPTMANTNOPL-----LEKTKOLRLDTD-----QV 455
DB 1372 SEKYIKKIBGSLSECKSKINSTLDKIDECIKNINVLKNNIINTEINTYXXFNAAE 1431
QY 456 YGNATYNEENGRVVDGNSMSEVLPOIOETTARILFENGKDLNVERRIAANPSDPLE 515
DB 1432 YNKIVLSNENN---TEWADNKSQYTLTKRNNGT---NDHDYNIKEIK-SHKDKSNGTK 1483
QY 516 TTKPDMLKEALKIAFGNEPNCNLOYGKD-----TTEPDFNPDQOTSQNIKNQ 565
DB 1484 -TEADOKKKAIOK-----NXELPEQYKEEVLLNKKYAVELKNKFD-KTKNDSKOI 1533
QY 566 LAELNATNIYVLDKIKILNKKM-----ILIRDRPHYDRNNIIVAGDESVYKAEHREV 619
DB 1534 IKEIKDANHCTLESCKSEKKMEIKNEKIHIEDEVANDKNSKMAITSIKVSEPEPKTI 1593
QY 620 IN-----STFEGILLMIDKIRIKILSGYIETED-----BELKVINIDRYDMLNS 666
DB 1594 IKINEITSDDCKLENTDLEKOISMLSDQETKLENGKOLKLEEL-----LE 1644
QY 667 SLRODGTFLIDKXKNDKPLYSN-----PARKYVAVYAKENTJINSNGDTS 717
DB 1645 SLKOKNIEDOKKLEDEVNSKIKNIEMTVBQHKNIETG-I-VEKINELAKTKNKKXIES 1702
QY 718 TNGIKK 723
DB 1703 TKELIK 1708

DT 01-MAY-1999 (Trembl) 10, (Created)
DT 01-MAY-1999 (Trembl) 10, (Last sequence update)
DT 01-DEC-2001 (Trembl) 19, (Last annotation update)
DE LARGE SUPERNATANT PROTEIN 1.
GN LSPAL
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales; Haemophilus.
OX NCBI_Taxid=730;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=3500;
RX MEDLINE=9030326; PubMed=9811662;
RA Ward C.R., Lumley S.R., Lattimer J.L., Cope L.D., Hansen E.J.;
RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein."
RL J. Bacteriol. 180:6013-6022(1998).
DR EMBL: AF057695; AAC79757.1; -
DR InterPro: IPR000130; Zn_MTPepidase.
DR SMART: SM00235; Zmnc; 1.
SQ SEQUENCE 4152 AA; 456173 MW; 7082DEDC988AB9F3 CRC64;

Query Match 4.8%; Score 183; DB 2; Length 4152;
Best Local Similarity 22.2%; Pred. No. 1.6; 321; Indels 218; Gaps 45;
Matches 187; Conservative 118; Mismatches

QY 10 NESESSGILGTYFSDLANFO-APVYVTSSTGPD-----LSPSELENIPSENGYF 60
DB 162 SRESSTQ-IVGKLHANILOKREKILINQYTDHESNIGALVPAKKADLIYV--- 216
QY 61 QSAIISGPIKAKSDERTPATSDNHYTAWDQEVINKASN-----SKIKLEGRLY 114
DB 217 PNCITLNGKYNTINIREVYTS-----DILPHREGLISVRKRYTIDRGVYA 264
QY 115 QIKIQY-----QRENTEKGLDFLYWDSQNKKEVISSD-----NQLPELKQSS 161
DB 265 TNGLSFEVYARNIDQNG--KIVAKTENOKSVNPANITTPAOSLNTKTEATPISS 321
QY 162 NSKKRSTAGPTVPDRNDGIPDSLEVEGYTVD---VKNKRTSEPMISNIHEKGLT 217
DB 322 GTSR--TSDTPAI-SADSAGSMYSGNIRFYVTDGAGVKNHGIIIFENDINIKMDGNA 377
QY 218 KYKSPKMSASDPYSDPEKYTGRIIDKNVSPAPRPLVAAPPIYHDMENIILSKNEDQ 277
DB 378 SLKEIYAK-----KDIDILAKDIELEKQLOANNKIIILNSTGKIMLRNASEV 425
QY 278 STONTSETRTIS-KATSTSRHTSEVHNAEVHASFPGIGSASGFSNSNSVAIDH 336
DB 426 SADNVKVSKEMLALENSMSANSGLVYIKLEVNRS-----SKVSAGTANIKRASNITIDG 480
QY 337 SLSLAGERTW-----AETMGLNTADTARLN-----ANIRYVN---T 369
DB 481 SSVAVANKITLAVTNNATLNNQSKLSAKIMELNTHTLNTSKLSQAKNITETLNLN 540
QY 370 GTAPIYNYLPTSLYLOK-----NOTLA-----TIKAKENO-LSQIAPNN-- 409
DB 541 GEA-----SLVAKEDINAKITNNGTITAGITANTIKALEMDNADALIAQNLN 591
QY 410 -----Y-----PSKNLAPALMAODSS-----TPITMANYNOFLEKTKOL 448
DB 592 FTYNGSHRYNKGDIYSKDAIYTESNSDSKLYDAONNLTVANNENTGOSSEI 651
QY 449 RLDTQVYGNI---ATYNEEN-GRVRYDTSNMSSEVLPOIOETTARILFENGKDLNVER- 503
DB 652 -----LHGNTVTLAKGNFTNSGNLTJTKRELNTSNIESTI---NAGNLTTGKRLLEVHSNT 702
QY 504 -----RTAAVNSDPLETTKPDMLT--KEALKIAGF---FNEPNGNLOYOGADITF 550
DB 703 TVANDGKLVSTENLNTSSTDTFTNGCTLLGLEALKIAGSGNFTVANSGLA--SNKSLDIY 761
QY 551 DFNF--DOOTSQNIKNQLAELNATNIYVLDK--IKLNAMNILLRDRREHYDRNNIYVGA 607

```

Db 762 GNETNGTIESYK-----LNTNNTFTFNATIKSYGVNLNTSQ-NFTNDSNGTYMSH 816
OY 608 D-ESYVKEAHEVYNSST-----BGLL-----NIDKDKILSGYIVEIEDTEGLKEVIND 658
Db 817 DLNLTQAN--LTKNKLACGGGLNLTAKGNITNDSNSTALAVLHSDNIN-----LNA 869
OY 659 RYDMLNSSL-RDQKTFIDPKYNDKLPYLSNPNKYVNVAVTKENTTINPSENGDTS 717
Db 870 NKKYNNIGETYSOAGNISVEAKLHNDVKL-----SGNITTTTK-----SGNATVK 915
OY 718 TNGI 721
Db 916 TNSI 919

```

Search completed: October 4, 2002, 14:40:05
 Job time: 280 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 14:34:15 ; Search time 14.66 Seconds
(without alignments)
1226.278 Million cell updates/sec

Title: US-09-848-909-21
Perfect score: 1 EVKQENRLNSESSESSGGL.....STNGIKLILPSKKYETGZ 736
Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 2442594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCPTS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3774	99.9	735	1	US-08-021-601-4
2	3774	99.9	735	1	US-08-082-843B-4
3	3774	99.9	735	1	PCT-US94-01624-4
4	3729.5	98.7	903	1	US-08-021-601-12
5	3729.5	98.7	903	1	US-08-082-843B-12
6	3729.5	98.7	903	1	PCT-US94-01624-12
7	3597	95.2	719	1	US-08-082-843B-31
8	3597	95.2	719	5	PCT-US94-01624-31
9	1275	33.7	288	4	US-09-273-839A-8
10	792	21.0	881	4	US-08-960-780-32
11	792	21.0	881	4	US-09-073-899-32
12	790	20.9	884	1	US-08-471-043-5
13	790	20.9	884	2	US-08-471-043-5
14	790	20.9	884	2	US-08-463-483A-5
15	790	20.9	884	2	US-08-471-046A-5
16	790	20.9	884	2	US-08-470-566B-5
17	790	20.9	884	2	US-08-469-334-5
18	790	20.9	884	3	US-09-300-529-5
19	790	20.9	1346	1	US-08-471-043-23
20	790	20.9	1346	2	US-08-471-044-23
21	790	20.9	1346	2	US-08-463-483A-23
22	790	20.9	1346	2	US-08-471-046A-23
23	790	20.9	1346	2	US-08-470-566B-23
24	790	20.9	1346	2	US-08-469-334-23
25	790	20.9	1346	3	US-09-300-529-23
26	789	20.9	852	1	US-08-471-043-36
27	789	20.9	852	2	US-08-471-044-36

28	789	20.9	852	2	US-08-463-483A-36	Sequence 36, Appl
29	789	20.9	852	2	US-08-471-046A-36	Sequence 36, Appl
30	789	20.9	852	2	US-08-470-566B-36	Sequence 36, Appl
31	789	20.9	852	2	US-08-469-334-36	Sequence 36, Appl
32	789	20.9	852	3	US-09-300-529-36	Sequence 50, Appl
33	789	20.9	1338	1	US-08-471-043-50	Sequence 50, Appl
34	789	20.9	1338	2	US-08-471-044-50	Sequence 50, Appl
35	789	20.9	1338	2	US-08-463-483A-50	Sequence 50, Appl
36	789	20.9	1338	2	US-08-470-566B-50	Sequence 50, Appl
37	789	20.9	1338	2	US-08-470-566B-50	Sequence 50, Appl
38	789	20.9	1338	2	US-08-469-334-50	Sequence 50, Appl
39	789	20.9	1338	4	US-09-300-529-50	Sequence 50, Appl
40	775.5	20.5	784	4	US-09-371-013A-7	Sequence 7, Appl
41	756.5	20.0	834	1	US-08-471-043-21	Sequence 21, Appl
42	756.5	20.0	834	2	US-08-471-044-21	Sequence 21, Appl
43	756.5	20.0	834	2	US-08-463-483A-21	Sequence 21, Appl
44	756.5	20.0	834	2	US-08-470-566B-21	Sequence 21, Appl
45	756.5	20.0	834	2	US-08-470-566B-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-021-601-4
Sequence 4, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klmpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Alora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Protein Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/021.601
APPLICATION NUMBER: US/08/021.601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
MOLECULE TYPE: protein
US-08-021-601-4
Query Match 99.9%; Score 3774; DB 1; length 735;
Best Local Similarity 100.0%; Pred. No. 1.9e-258;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1. EVKQENRLNSESSESSGGLGYFSDLPQAPMVYTSSTGDLSPSELENIPEENQVF 60

```

Db      1 EVKQERRLNLSSESSSGGLGYFSDLNFOAPWVYSSSTGDLSPSELENIPSENOYF 60
QY      61 QSAIWSGFIKAKKDEYTPATSDADNHVTMMVDDQEVINKASNSKRLKRGRLYQIKQY 120
Db      61 QSAIWSGFIKAKKDEYTPATSDADNHVTMMVDDQEVINKASNSKRLKRGRLYQIKQY 120
QY      121 QRENPEKGLDFKLYMTDSQNKREYISSDNLOJPELKOKSSNSKRRKSTASGPTYPDDN 180
Db      121 QRENPEKGLDFKLYMTDSQNKREYISSDNLOJPELKOKSSNSKRRKSTASGPTYPDDN 180
QY      181 DGIPDSLEVEGYTVDAVKNKRTFLSPWISNIHEKKGLTKYKSSPEKMSSTADSPSEKVT 240
Db      181 DGIPDSLEVEGYTVDAVKNKRTFLSPWISNIHEKKGLTKYKSSPEKMSSTADSPSEKVT 240
QY      241 GRIDKNVSPKARHPLVAAYPIVHVMENIILSKNEOSTONTDESETTISKNTSRTHT 300
Db      241 GRIDKNVSPKARHPLVAAYPIVHVMENIILSKNEOSTONTDESETTISKNTSRTHT 300
QY      301 SEVHGAELVHASFDFDGSVSAGFSNSNSTVAIDHSLSLAGEETVAETMGINTADTARL 360
Db      301 SEVHGAELVHASFDFDGSVSAGFSNSNSTVAIDHSLSLAGEETVAETMGINTADTARL 360
QY      361 NANIRVNTGAPLYNVLPFTSLVIGKNQTLATIKAKENQSLQILAPNNYPSKMLAPIA 420
Db      361 NANIRVNTGAPLYNVLPFTSLVIGKNQTLATIKAKENQSLQILAPNNYPSKMLAPIA 420
QY      421 LNAODESPTITNNYNOFLELEKTKOLRLDTDOVYGNATYFENGVRVADTGSNMSEV 480
Db      421 LNAODESPTITNNYNOFLELEKTKOLRLDTDOVYGNATYFENGVRVADTGSNMSEV 480
QY      481 LPQIOETARIRIFNGKDLNIVERRIAAVNSDPLETTKPDMTLKEALKIAGFENPNL 540
Db      481 LPQIOETARIRIFNGKDLNIVERRIAAVNSDPLETTKPDMTLKEALKIAGFENPNL 540
QY      541 OYQGRITPEPFNDQOOTSQNKNOELAEANTNITVLDKIKLAKANNIILIRKRFPHDR 600
Db      541 OYQGRITPEPFNDQOOTSQNKNOELAEANTNITVLDKIKLAKANNIILIRKRFPHDR 600
QY      601 NNIAVGADESVKRAHREYINSTEGILLNDIRKILSGYIEIDEGTGLKEYINDRY 660
Db      601 NNIAVGADESVKRAHREYINSTEGILLNDIRKILSGYIEIDEGTGLKEYINDRY 660
QY      661 DMNINISLROGKTFIDFKYNDKLPYISNPKYKYNVAYATKENTLIINSENGDSTNG 720
Db      661 DMNINISLROGKTFIDFKYNDKLPYISNPKYKYNVAYATKENTLIINSENGDSTNG 720
QY      721 IKKILFSSKGYEIG 735
Db      721 IKKILFSSKGYEIG 735

RESULT 2
US-08-082-849B-4
; Sequence 4, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kimpel, Kurt R.
; APPLICANT: Arota, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

```

```

; MEDIAN TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-4

```

```

Query Match          99.9%; Score 3774; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 1.9e-258;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 EVKQERRLNLSSESSSGGLGYFSDLNFOAPWVYSSSTGDLSPSELENIPSENOYF 60
Db      1 EVKQERRLNLSSESSSGGLGYFSDLNFOAPWVYSSSTGDLSPSELENIPSENOYF 60
QY      61 QSAIWSGFIKAKKDEYTPATSDADNHVTMMVDDQEVINKASNSKRLKRGRLYQIKQY 120
Db      61 QSAIWSGFIKAKKDEYTPATSDADNHVTMMVDDQEVINKASNSKRLKRGRLYQIKQY 120
QY      121 QRENPEKGLDFKLYMTDSQNKREYISSDNLOJPELKOKSSNSKRRKSTASGPTYPDDN 180
Db      121 QRENPEKGLDFKLYMTDSQNKREYISSDNLOJPELKOKSSNSKRRKSTASGPTYPDDN 180
QY      181 DGIPDSLEVEGYTVDAVKNKRTFLSPWISNIHEKKGLTKYKSSPEKMSSTADSPSEKVT 240
Db      181 DGIPDSLEVEGYTVDAVKNKRTFLSPWISNIHEKKGLTKYKSSPEKMSSTADSPSEKVT 240
QY      241 GRIDKNVSPKARHPLVAAYPIVHVMENIILSKNEOSTONTDESETTISKNTSRTHT 300
Db      241 GRIDKNVSPKARHPLVAAYPIVHVMENIILSKNEOSTONTDESETTISKNTSRTHT 300
QY      301 SEVHGAELVHASFDFDGSVSAGFSNSNSTVAIDHSLSLAGEETVAETMGINTADTARL 360
Db      301 SEVHGAELVHASFDFDGSVSAGFSNSNSTVAIDHSLSLAGEETVAETMGINTADTARL 360
QY      361 NANIRVNTGAPLYNVLPFTSLVIGKNQTLATIKAKENQSLQILAPNNYPSKMLAPIA 420
Db      361 NANIRVNTGAPLYNVLPFTSLVIGKNQTLATIKAKENQSLQILAPNNYPSKMLAPIA 420
QY      421 LNAODESPTITNNYNOFLELEKTKOLRLDTDOVYGNATYFENGVRVADTGSNMSEV 480
Db      421 LNAODESPTITNNYNOFLELEKTKOLRLDTDOVYGNATYFENGVRVADTGSNMSEV 480
QY      481 LPQIOETARIRIFNGKDLNIVERRIAAVNSDPLETTKPDMTLKEALKIAGFENPNL 540
Db      481 LPQIOETARIRIFNGKDLNIVERRIAAVNSDPLETTKPDMTLKEALKIAGFENPNL 540
QY      541 OYQGRITPEPFNDQOOTSQNKNOELAEANTNITVLDKIKLAKANNIILIRKRFPHDR 600
Db      541 OYQGRITPEPFNDQOOTSQNKNOELAEANTNITVLDKIKLAKANNIILIRKRFPHDR 600
QY      601 NNIAVGADESVKRAHREYINSTEGILLNDIRKILSGYIEIDEGTGLKEYINDRY 660
Db      601 NNIAVGADESVKRAHREYINSTEGILLNDIRKILSGYIEIDEGTGLKEYINDRY 660

```

DB 601 NNIAVGADESVYKKAHREVINSSTGGLLNDKIDKIKLLISGYIVLEDEGLKEVINDRY 660
QY 661 DMLNSSLRODKGTFIDFKRYNDKILPLYSNPNKYVAVYATKENTINPNSNGTSTNG 720
DB 661 DMLNSSLRODKGTFIDFKRYNDKILPLYSNPNKYVAVYATKENTINPNSNGTSTNG 720
QY 721 IKKILFSSKGYEIG 735
DB 721 IKKILFSSKGYEIG 735

RESULT 3
PCT-US94-01624-4
Sequence 4, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Aroca, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-4

Query Match 99.9% Score 3774: DB 5: Length 735;
Best Local Similarity 100.0% Pctd No 1.9e-258;
Matches 735: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVKQENRLNLESSSSQGLITFYPSDLNFOAPVVTSTTGDI STPSSSELENISEKQVF 60
DB 1 EVKQENRLNLESSSSQGLITFYPSDLNFOAPVVTSTTGDI STPSSSELENISEKQVF 60
QY 61 OSAINSGFIYKKSDEYTFATSDNHNVTMVDDEVINKASNSNKTREKGLQVQIKIY 120
DB 61 OSAINSGFIYKKSDEYTFATSDNHNVTMVDDEVINKASNSNKTREKGLQVQIKIY 120
QY 121 ORENPTFEGDGLFKLYTDSQNKKEVTSNDLQPELKOKSSNSRKRRTSAGPTVPDRDN 180
DB 121 ORENPTFEGDGLFKLYTDSQNKKEVTSNDLQPELKOKSSNSRKRRTSAGPTVPDRDN 180
QY 181 DGIPLSLVEGEYIVDVKNKRTFLSPWISNIHKKGLTKYKSSPEKWSSTASDPYSDPEKYV 240

DB 181 DGIPLSLVEGEYIVDVKNKRTFLSPWISNIHKKGLTKYKSSPEKWSSTASDPYSDPEKYV 240
QY 241 GRIDKNVSPPEARHPLVAAPYIVHVDMENILSKNEQOSTONTQSETRTISKNTSTSRTH 300
DB 241 GRIDKNVSPPEARHPLVAAPYIVHVDMENILSKNEQOSTONTQSETRTISKNTSTSRTH 300
QY 301 SEVHGNAVHASFEDIGGSVSAAGFSNSNSTVAIDHSLSLAGEFNTAEMTGINTADTARL 360
DB 301 SEVHGNAVHASFEDIGGSVSAAGFSNSNSTVAIDHSLSLAGEFNTAEMTGINTADTARL 360
QY 361 NANIRVYNTGAPTYVNLPTSLVYGRKQTLATIKAKENOLSOIILAVNNYPSKMLAPIA 420
DB 361 NANIRVYNTGAPTYVNLPTSLVYGRKQTLATIKAKENOLSOIILAVNNYPSKMLAPIA 420
QY 421 LNKODFSSPTNNYNOLELEKTKOLRLTDQVGNLATYNSPENSERVAVDTGSNMGSEV 480
DB 421 LNKODFSSPTNNYNOLELEKTKOLRLTDQVGNLATYNSPENSERVAVDTGSNMGSEV 480
QY 481 LPOIOTTRATIFNGKDLNVERRIANVPSDPLETTKPDMTIKKALKAIFGNPNNGNL 540
DB 481 LPOIOTTRATIFNGKDLNVERRIANVPSDPLETTKPDMTIKKALKAIFGNPNNGNL 540
QY 541 QYQKDIETFDNFDDQTSQNIKOLAELNATNTIYLDKIKLVKENTILIPKRPYHPR 600
DB 541 QYQKDIETFDNFDDQTSQNIKOLAELNATNTIYLDKIKLVKENTILIPKRPYHPR 600
QY 601 NNIAVGADESVYKKAHREVINSSTGGLLNDKIDKIKLLISGYIVLEDEGLKEVINDRY 660
DB 601 NNIAVGADESVYKKAHREVINSSTGGLLNDKIDKIKLLISGYIVLEDEGLKEVINDRY 660
QY 661 DMLNSSLRODKGTFIDFKRYNDKILPLYSNPNKYVAVYATKENTINPNSNGTSTNG 720
DB 661 DMLNSSLRODKGTFIDFKRYNDKILPLYSNPNKYVAVYATKENTINPNSNGTSTNG 720
QY 721 IKKILFSSKGYEIG 735
DB 721 IKKILFSSKGYEIG 735

RESULT 4
US-08-021-601-12
Sequence 12, Application US/08021601
Patient No 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Aroca, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Owendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 903 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-021-601-12

Query Match 98.7%; Score 3729.5; DB 1; Length 903;
 Best Local Similarity 99.2%; Pred. No. 3, 9e-255;
 Matches 728; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

```

QY 1 EVKQENRLNSESSESSGGLGYFSDLNFOAPMVVTSSTTGLSPSELENIPESENOYF 60
DB 1 EVKQENRLNSESSESSGGLGYFSDLNFOAPMVVTSSTTGLSPSELENIPESENOYF 60
QY 61 QSAIWSGFIKVKKSDEYTFATSADNHVYTMWVDDQEVYINKASNSKRIKLEGRLYQIKIY 120
DB 61 QSAIWSGFIKVKKSDEYTFATSADNHVYTMWVDDQEVYINKASNSKRIKLEGRLYQIKIY 120
QY 121 QRENPTKGLDFKLYMTDSQNKKEVYISSDNLOPELKOKSSNSRKKSTAGPTVYDRDN 180
DB 121 QRENPTKGLDFKLYMTDSQNKKEVYISSDNLOPELKOKSSNSRKKSTAGPTVYDRDN 180
QY 181 DGIPDSELEVEGYTVDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSADPYDPEKYT 240
DB 181 DGIPDSELEVEGYTVDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSADPYDPEKYT 240
QY 241 GRIDKIVSPEARHPLVAAPYIVHVMENIILSKNEQSTQNTDSEFTTISKNTSRTHT 300
DB 241 GRIDKIVSPEARHPLVAAPYIVHVMENIILSKNEQSTQNTDSEFTTISKNTSRTHT 300
QY 301 SEVHGAENVAEVSFDIGGSVAGFSNSNSTVAIDHSLSLAGERTVAETMGLNTADTARL 360
DB 301 SEVHGAENVAEVSFDIGGSVAGFSNSNSTVAIDHSLSLAGERTVAETMGLNTADTARL 360
QY 361 NANIRYVNTGTAPIYVNLPTTSLVYGNKOTLATIKAKENQSLQILAPNNYPSKNIAPIA 420
DB 361 NANIRYVNTGTAPIYVNLPTTSLVYGNKOTLATIKAKENQSLQILAPNNYPSKNIAPIA 420
QY 421 LNNODFSSPTITNNYNOFLEKTKOLRLDPOYGNATATNPNKRVVPOGSMSEY 480
DB 421 LNNODFSSPTITNNYNOFLEKTKOLRLDPOYGNATATNPNKRVVPOGSMSEY 480
QY 481 LPOIETTKARIIFNGKDLNVERRIAAVNSDPLETTKPDMTLAKALKIAPGNEPENG 540
DB 481 LPOIETTKARIIFNGKDLNVERRIAAVNSDPLETTKPDMTLAKALKIAPGNEPENG 540
QY 541 OYQKQITTEFDPNFDQTSQNIKNOLAEANNTIYVLDKIKLAKNNMILLIDKREPHDR 600
DB 541 OYQKQITTEFDPNFDQTSQNIKNOLAEANNTIYVLDKIKLAKNNMILLIDKREPHDR 600
QY 601 NNIAVADSESVKKAHREVIINSSTBGLLNDKDKIRKISGTYVEIEDTEGLAKEVINDRY 660
DB 601 NNIAVADSESVKKAHREVIINSSTBGLLNDKDKIRKISGTYVEIEDTEGLAKEVINDRY 660
QY 661 DMLNSSLRODGKTFIDFKKYNDKPLIISPNKXVYVAATKENTJINSNGDSTING 720
DB 661 DMLNSSLRODGKTFIDFKKYNDKPLIISPNKXVYVAATKENTJINSNGDSTING 720
QY 721 IKRIL---IFSKKG 731
DB 721 IKRILKKVYLAKKG 734

```

GENERAL INFORMATION:
 APPLICANT: Leppla, Stephen H.
 APPLICANT: Klimpel, Kurt R.
 APPLICANT: Arota, Naveen
 APPLICANT: Singh, Yogendra
 APPLICANT: Nichols, Peter J.
 TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
 TITLE OF INVENTION: Related Methods
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/082,849B
 FILING DATE: 25-JUN-1993
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/021,601
 FILING DATE: 12-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Weber, Kenneth A.
 REGISTRATION NUMBER: 31,677
 REFERENCE/DOCKET NUMBER: 15280-161-1
 TELEPHONE: (415) 576-0300
 TELEFAX: (415) 576-0200
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 903 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-082-849B-12

Query Match 98.7%; Score 3729.5; DB 1; Length 903;
 Best Local Similarity 99.2%; Pred. No. 3, 9e-255;
 Matches 728; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

```

QY 1 EVKQENRLNSESSESSGGLGYFSDLNFOAPMVVTSSTTGLSPSELENIPESENOYF 60
DB 1 EVKQENRLNSESSESSGGLGYFSDLNFOAPMVVTSSTTGLSPSELENIPESENOYF 60
QY 61 QSAIWSGFIKVKKSDEYTFATSADNHVYTMWVDDQEVYINKASNSKRIKLEGRLYQIKIY 120
DB 61 QSAIWSGFIKVKKSDEYTFATSADNHVYTMWVDDQEVYINKASNSKRIKLEGRLYQIKIY 120
QY 121 QRENPTKGLDFKLYMTDSQNKKEVYISSDNLOPELKOKSSNSRKKSTAGPTVYDRDN 180
DB 121 QRENPTKGLDFKLYMTDSQNKKEVYISSDNLOPELKOKSSNSRKKSTAGPTVYDRDN 180
QY 181 DGIPDSELEVEGYTVDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSADPYDPEKYT 240
DB 181 DGIPDSELEVEGYTVDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSADPYDPEKYT 240
QY 241 GRIDKIVSPEARHPLVAAPYIVHVMENIILSKNEQSTQNTDSEFTTISKNTSRTHT 300
DB 241 GRIDKIVSPEARHPLVAAPYIVHVMENIILSKNEQSTQNTDSEFTTISKNTSRTHT 300
QY 301 SEVHGAENVAEVSFDIGGSVAGFSNSNSTVAIDHSLSLAGERTVAETMGLNTADTARL 360
DB 301 SEVHGAENVAEVSFDIGGSVAGFSNSNSTVAIDHSLSLAGERTVAETMGLNTADTARL 360
QY 361 NANIRYVNTGTAPIYVNLPTTSLVYGNKOTLATIKAKENQSLQILAPNNYPSKNIAPIA 420

```

RESULT 5
 US-08-082-849B-12
 Sequence 12, Application US/08082849B
 Patent No. 5677274

```

|||||
Db 361 NANIRVNTGAPLYNVLPPTSLVYGNKQTLATIKAKENOLSOILAPNNYPSKMLAFIA 420
Oy 421 LNAODFSSPTITMNTNOFLEKTKOLRLDQVGNATATYENNGRVAVDTGSNMSEV 480
Db 421 LNAODFSSPTITMNTNOFLEKTKOLRLDQVGNATATYENNGRVAVDTGSNMSEV 480
Oy 481 LPQIETTARITIFNGKDLNVERRIAANVPSDPLETTKPDWTLKALKAFAFGNPNL 540
Db 481 LPQIETTARITIFNGKDLNVERRIAANVPSDPLETTKPDWTLKALKAFAFGNPNL 540
Oy 541 OYOGKDITEFPNPDQTSNKNKQLAELNATNTYVLDKIKLAKNNLILDRKFFHDR 600
Db 541 OYOGKDITEFPNPDQTSNKNKQLAELNATNTYVLDKIKLAKNNLILDRKFFHDR 600
Oy 601 NNINAGADESVYKAAHREYVNSTGGLLNIDDKILSGYVEIEDTGBGLKEYINDRY 660
Db 601 NNINAGADESVYKAAHREYVNSTGGLLNIDDKILSGYVEIEDTGBGLKEYINDRY 660
|||||
DQKTFIDFKYNDKLPYLISNPNYKVNVAATKENTLINSENGDSTNG 720
DQKTFIDFKYNDKLPYLISNPNYKVNVAATKENTLINSENGDSTNG 720
SRKG 731
GKKG 734
cation PC/TUS9401624
ON:
pla, Stephen H.
mpel, Kurt R.
ra, Naveen
gh, Yogendra
hols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSER: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-12

```

```

Query Match 98.7%; Score 3729.5; DB 5; Length 903;
Best Local Similarity 99.2%; Pred. No. 3,5e-255;
Matches 728; Conservative 1; Mismatches 2; Indels 3; Gaps 1;
Oy 1 EYKOENRLNNESSSSOGILGYPSDLNFOAPMVVTSSTTGDLSIPSELENIPESENOYF 60
Db 1 EYKOENRLNNESSSSOGILGYPSDLNFOAPMVVTSSTTGDLSIPSELENIPESENOYF 60
Oy 61 OSAINSGFIKKYKSDYETPATSDNHYTMVDDQEVYKASNSNKRLKLEKGLYKIQY 120
Db 61 OSAINSGFIKKYKSDYETPATSDNHYTMVDDQEVYKASNSNKRLKLEKGLYKIQY 120
Oy 121 ORENPTKGLDFTKYWTDSQNKKEYISSDLQLPELKOKSSNKKRSTASAGTPPDDN 180
Db 121 ORENPTKGLDFTKYWTDSQNKKEYISSDLQLPELKOKSSNKKRSTASAGTPPDDN 180
Oy 181 DGIPSLVEBGTIVDYKAKRTFLSPWISNHEKKGLTKYKSSPKASSTASDPSEFYV 240
Db 181 DGIPSLVEBGTIVDYKAKRTFLSPWISNHEKKGLTKYKSSPKASSTASDPSEFYV 240
Oy 241 GRIDKNVSEARHPLVAATPIVHDMENILSKNEDOSTONTDSERTISKNTSTSRHT 300
Db 241 GRIDKNVSEARHPLVAATPIVHDMENILSKNEDOSTONTDSERTISKNTSTSRHT 300
Oy 301 SEVHGNAEVHASFEDIGGSVSGFSNSNSTVAIDHSLSLAGEEVAETMGINTADTARL 360
Db 301 SEVHGNAEVHASFEDIGGSVSGFSNSNSTVAIDHSLSLAGEEVAETMGINTADTARL 360
Oy 361 NANIRVNTGAPLYNVLPPTSLVYGNKQTLATIKAKENOLSOILAPNNYPSKMLAFIA 420
Db 361 NANIRVNTGAPLYNVLPPTSLVYGNKQTLATIKAKENOLSOILAPNNYPSKMLAFIA 420
Oy 421 LNAODFSSPTITMNTNOFLEKTKOLRLDQVGNATATYENNGRVAVDTGSNMSEV 480
Db 421 LNAODFSSPTITMNTNOFLEKTKOLRLDQVGNATATYENNGRVAVDTGSNMSEV 480
Oy 481 LPQIETTARITIFNGKDLNVERRIAANVPSDPLETTKPDWTLKALKAFAFGNPNL 540
Db 481 LPQIETTARITIFNGKDLNVERRIAANVPSDPLETTKPDWTLKALKAFAFGNPNL 540
Oy 541 OYOGKDITEFPNPDQTSNKNKQLAELNATNTYVLDKIKLAKNNLILDRKFFHDR 600
Db 541 OYOGKDITEFPNPDQTSNKNKQLAELNATNTYVLDKIKLAKNNLILDRKFFHDR 600
Oy 601 NNINAGADESVYKAAHREYVNSTGGLLNIDDKILSGYVEIEDTGBGLKEYINDRY 660
Db 601 NNINAGADESVYKAAHREYVNSTGGLLNIDDKILSGYVEIEDTGBGLKEYINDRY 660
Oy 661 DMLNSSLRQDKTFIDFKYNDKLPYLISNPNYKVNVAATKENTLINSENGDSTNG 720
Db 661 DMLNSSLRQDKTFIDFKYNDKLPYLISNPNYKVNVAATKENTLINSENGDSTNG 720
Oy 721 IKKTL--IFSKG 731
Db 721 IKKTLKKVYLGKKG 734
RESULT 7
US-08-849B-31
Sequence 31, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klompel, Kurt R.
APPLICANT: Aroa, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor

```

```

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-31

```

```

Query Match          95.2%; Score 3597; DB 1; Length 719;
Best Local Similarity 95.5%; Pred. No. 5.7e-246;
Matches 706; Conservative 4; Mismatches 5; Indels 24; Gaps 2;

OY 1 EVKQENRLNLESESSSQGLGYFSDLNFOAPVYVTSSTGDLSPSSSELENIPSENOYF 60
DB 1 EVKQENRLNLESESSSQGLGYFSDLNFOAPVYVTSSTGDLSPSSSELENIPSENOYF 60
OY 61 QSAIWSGFIKKKSDYEYFATSADNHVTMMVWVDOEYINKASNSNKRIRLEKGLYOIKIOY 120
DB 61 QSAIWSGFIKKKSDYEYFATSADNHVTMMVWVDOEYINKASNSNKRIRLEKGLYOIKIOY 120
OY 121 QRENPEKGLDEKLYWDSQNKKEVYSSDNLOLPELKOKSSNS----RKRSTSAQPTVP 176
DB 121 QRENPEKGLDEKLYWDSQNKKEVYSSDNLOLPELKOKSSNS----RKRSTSAQPTVP 180
OY 177 DRONDGIPSLSEVGGYVDVKNKRTFLSPWISNHEKGLTRYKSSPEKMSPTASPYGDF 236
DB 177 DRONDGIPSLSEVGGYVDVKNKRTFLSPWISNHEKGLTRYKSSPEKMSPTASPYGDF 240
OY 181 DRONDGIPSLSEVGGYVDVKNKRTFLSPWISNHEKGLTRYKSSPEKMSPTASPYGDF 240
DB 181 DRONDGIPSLSEVGGYVDVKNKRTFLSPWISNHEKGLTRYKSSPEKMSPTASPYGDF 240
OY 237 EKYTGILDKNVSPKARHPLVAAPIVHVDMENILSKNEDOSTOMTOSERTSKSTTS 296
DB 237 EKYTGILDKNVSPKARHPLVAAPIVHVDMENILSKNEDOSTOMTOSERTSKSTTS 300
OY 241 EKYTGILDKNVSPKARHPLVAAPIVHVDMENILSKNEDOSTOMTOSERTSKSTTS 300
DB 241 EKYTGILDKNVSPKARHPLVAAPIVHVDMENILSKNEDOSTOMTOSERTSKSTTS 300
OY 297 RHTHSVHNAVYHASFDFIGGSVSGFSNSNSTVAIDHSISLAGRTMAETMGLNTRAD 356
DB 297 RHTHSVHNAVYHASFDFIGGSVSGFSNSNSTVAIDHSISLAGRTMAETMGLNTRAD 360
OY 301 RHTHSVHNAVYHASFDFIGGSVSGFSNSNSTVAIDHSISLAGRTMAETMGLNTRAD 360
DB 301 RHTHSVHNAVYHASFDFIGGSVSGFSNSNSTVAIDHSISLAGRTMAETMGLNTRAD 360
OY 357 TARLANINRYVTGAPRIYVLPPTSILVCKNQTLATIRAKNQLSQILAPNNYPSKNL 416
DB 357 TARLANINRYVTGAPRIYVLPPTSILVCKNQTLATIRAKNQLSQILAPNNYPSKNL 420
OY 417 APILANADDESSPTITMANTNOFLELEKTKOLRLDVOYGINATYFNGRVRVDTGSN 476
DB 417 APILANADDESSPTITMANTNOFLELEKTKOLRLDVOYGINATYFNGRVRVDTGSN 480
OY 476 APILANADDESSPTITMANTNOFLELEKTKOLRLDVOYGINATYFNGRVRVDTGSN 480
DB 476 APILANADDESSPTITMANTNOFLELEKTKOLRLDVOYGINATYFNGRVRVDTGSN 480
OY 497 WSEVLQIOETTARITFNGKDLNVERRIAANVNSDPLETTPDMLTKEALKIAGFNEP 536
DB 497 WSEVLQIOETTARITFNGKDLNVERRIAANVNSDPLETTPDMLTKEALKIAGFNEP 540
OY 537 WSEVLQIOETTARITFNGKDLNVERRIAANVNSDPLETTPDMLTKEALKIAGFNEP 540
DB 537 WSEVLQIOETTARITFNGKDLNVERRIAANVNSDPLETTPDMLTKEALKIAGFNEP 540
OY 597 NGNLOYQGDITFEFNFDOOTSQNIKNQLAELNATNITYVLDKIKINAMNILLRDRF 596
DB 597 NGNLOYQGDITFEFNFDOOTSQNIKNQLAELNATNITYVLDKIKINAMNILLRDRF 596

```

```

DB 521 NGNLOYQGDITFEFNFDOOTSQNIKNQLAELNATNITYVLDKIKINAMNILLRDRF 580
OY 521 NGNLOYQGDITFEFNFDOOTSQNIKNQLAELNATNITYVLDKIKINAMNILLRDRF 580
DB 597 HYDRNNIAGADESVYKAEHREAVINSSTEGLLINDKIRKILSGTIVEIEDTGLKEVI 656
OY 597 HYDRNNIAGADESVYKAEHREAVINSSTEGLLINDKIRKILSGTIVEIEDTGLKEVI 660
DB 581 HYDRNNIAGADESVYKAEHREAVINSSTEGLLINDKIRKILSGTIVEIEDTGLKEVI 640
OY 657 NDYDMLNLSLQDKRTYDEKRYNDKPLYSNPNTKVNYAVYKENTIIINPSENGDT 716
DB 641 NDYDMLNLSLQDKRTYDEKRYNDKPLYSNPNTKVNYAVYKENTIIINPSENGDT 700
OY 717 STNGIKRLILFSKKGEIG 735
DB 701 STNGIKRLILFSKKGEIG 719

```

```

RESULT 8
PCT-US94-01624-31
Sequence 31, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klmpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSER: TOWNSEND AND TOWNSEND KHOURIE and GREEN
STREET: Steuart Street Tower, 20th floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-31

```

```

Query Match          95.2%; Score 3597; DB 5; Length 719;
Best Local Similarity 95.5%; Pred. No. 5.7e-246;
Matches 706; Conservative 4; Mismatches 5; Indels 24; Gaps 2;

OY 1 EVKQENRLNLESESSSQGLGYFSDLNFOAPVYVTSSTGDLSPSSSELENIPSENOYF 60
DB 1 EVKQENRLNLESESSSQGLGYFSDLNFOAPVYVTSSTGDLSPSSSELENIPSENOYF 60
OY 61 QSAIWSGFIKKKSDYEYFATSADNHVTMMVWVDOEYINKASNSNKRIRLEKGLYOIKIOY 120
DB 61 QSAIWSGFIKKKSDYEYFATSADNHVTMMVWVDOEYINKASNSNKRIRLEKGLYOIKIOY 120

```

Fri Oct 4 14:31:03 2002

us-09-848-909-21.ra1

Page 7

121 QRENPTEKGLDFKLYTWSQNKKEVYSSDNLOPELKQSSNS-----RKRSFSGPTVP 176
121 QRENPTEKGLDFKLYTWSQNKKEVYSSDNLOPELKQSSNSMTATIMQGNFLOGPTVP 180
177 DRONDIDPDLSEVEGYTVQVKNKRTFLSPWISNHEKGLTKYKSPKSTSPDYDF 236
181 DRONDIDPDLSEVEGYTVQVKNKRTFLSPWISNHEKGLTKYKSPKSTSPDYDF 240
237 EKVGRIDKNVSPFARHPLVAAPPIVYVDMENILLSKNEDOSTONTSETRISKTSPS 296
241 EKVGRIDKNVSPFARHPLVAAPPIVYVDMENILLSKNEDOSTONTSETRISKTSPS 300
297 RHTSEVHGAAVHASFPGDGSVSGFSGNSSTVAIDHSLSLAGEPTMAETMGLNTAD 356
301 RHTSEVHGAAVHASFPGDGSVSGFSGNSSTVAIDHSLSLAGEPTMAETMGLNTAD 360
357 TARKLANIRVYNTGAPPIYVNLPTSLVYKGNQTLATKAKENOLSOILADPNYPSKML 416
361 TARKLANIRVYNTGAPPIYVNLPTSLVYKGNQTLATKAKENOLSOILADPNYPSKML 420
417 APIALNADDFSTPTMNTYNOPIELEKTKOLADTOGYGNATYFENGVRVDTGSN 476
421 APIALNADDFSTPTMNTYNOPIELEKTKOLADTOGYGNATYFENGVRVDTGSN 460
477 WSEVLPJOETFTARIENGKOLANVERIAVNSDPLETTKPDMLKFAKIAFGFNEP 536
461 WSEVLPJOETFTARIENGKOLANVERIAVNSDPLETTKPDMLKFAKIAFGFNEP 520
537 NGNLQYQCKDITEFDPNEDQTSNIKOLAELNATNTYVLDKIKLNKNNILIRDKRF 596
521 NGNLQYQCKDITEFDPNEDQTSNIKOLAELNATNTYVLDKIKLNKNNILIRDKRF 580
597 HYDRNNIAGADESVYKAEHREYINSSTEGILLNIDKIRKILSGYVEIEDEBLAKVI 656
581 HYDRNNIAGADESVYKAEHREYINSSTEGILLNIDKIRKILSGYVEIEDEBLAKVI 640
657 NDYEDMLNSSLRODGKTFIDEKYNKDLPLYSINPNYVYVNTYKENTINSENGDT 716
641 NDYEDMLNSSLRODGKTFIDEKYNKDLPLYSINPNYVYVNTYKENTINSENGDT 700
717 STNGIKKILIFSKGYEIG 735
701 STNGIKKILIFSKGYEIG 719
RESULT 9
US-09-273-839A-8
Sequence 8, Application US/09273839A
Patent No. 6329156
GENERAL INFORMATION:
APPLICANT: Citino, Nick M.
APPLICANT: Jackson, Bruce E.
APPLICANT: Lehmetz, Paul J.
TITLE OF INVENTION: Disruption of Anthrax Toxin Binding to Cell Surface
FILE REFERENCE: S-89/062
CURRENT APPLICATION NUMBER: US/09/273,839A
CURRENT FILING DATE: 1999-03-22
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 288
TYPE: PRT
ORGANISM: Bacillus anthracis
US-09-273-839A-8
Query Match 33.7%; Score 1275; DB 4; Length 288;
Best Local Similarity 100.0%; Freq. No. 1,36-82;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 486 ETVARIIFGKDLNIVERRIAVNSDPLETTKPDMLKFAKIAFGFNEPNCNLOYQCK 545
|||||

DB 29 ETVARIIFGKDLNIVERRIAVNSDPLETTKPDMLKFAKIAFGFNEPNCNLOYQCK 88
QY 546 DITEFDPNEDQTSNIKOLAELNATNTYVLDKIKLNKNNILIRDKRFHYDRNNIAY 605
DB 89 DITEFDPNEDQTSNIKOLAELNATNTYVLDKIKLNKNNILIRDKRFHYDRNNIAY 148
QY 606 GADESVEKAEHREYINSSTEGILLNIDKIRKILSGYVEIEDEBLAKVI 665
DB 149 GADESVEKAEHREYINSSTEGILLNIDKIRKILSGYVEIEDEBLAKVI 208
QY 666 SSLRODGKTFIDEKYNKDLPLYSINPNYVYVNTYKENTINSENGDTSTNGIKIL 725
DB 209 SSLRODGKTFIDEKYNKDLPLYSINPNYVYVNTYKENTINSENGDTSTNGIKIL 268
QY 726 IFSKGYEIG 735
DB 269 IFSKGYEIG 278
RESULT 10
US-08-960-780-32
Sequence 32, Application US/08960780
Patent No. 6204435
GENERAL INFORMATION:
APPLICANT: Feltelson, Jerald S.
APPLICANT: Schaeff, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeltz, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
TITLE OF INVENTION: No. 6204435el Pesticidal toxins and Nucleotide
TITLE OF INVENTION: Sequences Which Encode These Toxins
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,780
FILING DATE: 30-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA-708
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 881 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: peptide
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 177c8
US-08-960-780-32

Query Match	21.0%;	Score 792;	DB 4;	Length 881;
Best Local Similarity	30.4%;	Pred. No. 7,9e-48;		
Matches 246;	Conservative 135;	Mismatches 297;	Indels 130;	Gaps 34

[illegible]

APPLICANT: Dullum, Charles Joseph
 APPLICANT: Muller-Cohn, Judy
 APPLICANT: Stamp, Lisa
 APPLICANT: Morrill, George
 APPLICANT: Flinstad-lee, Stacey
 TITLE OF INVENTION: No. 6242669d Pesticidal Toxins and Nucleotide
 TITLE OF INVENTION: Sequences Which Encode These Toxins
 NUMBER OF SEQUENCES: 144
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
 STREET: 2421 N.W. 41st Street, Suite A-1
 CITY: Gainesville
 STATE: FL
 COUNTRY: US
 ZIP: 32606-6669
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/073,898
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/029,848
 FILING DATE: 30-Oct-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/960,760
 FILING DATE: 30-Oct-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Sanders, Jay M.
 REGISTRATION NUMBER: 39,355
 REFERENCE/DOCKET NUMBER: MA-708C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 352-375-8100
 TELEFAX: 352-372-5800
 INFORMATION FOR SEQ ID NO: 37:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 881 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: PSI177C8

Query Match	21.08%	Score 792	DB 48	Length 881	
Best local similarity	70.48%	Pred. No. 7	9e-48		
Matches 246:	Conservative 135	Mismatches 297	Indels 130	Gaps 34	

QY	4	QENRLINSSSSSSGGLGYPSDLNFAQAPVYVTSSTTGGDLSFISAELEN--IISSEKQFO	61
DB	39	QKRNQ--QKEMDRKGLGYRGKDP--SULMEAFYIDGSLITQYANKLKKQKQEO	94
QY	62	SAIWSGFIKYKSDSEYTPRATSDNHTYMMVQDQYIKKASMSKRLKRGKVIQIKQO	121
DB	95	STRMGLQKQSEYDDFTYMLSEDDALINCKIKISKKRGEKQVYHLERGLVPIKIEYO	154
QY	122	REMPKPF--FKIYWDSONKKEVYSNDQLPRLKQKS--	N 162
DB	155	SD--FKNPIDSKTFKELKFLKIFKIDSONPOVOQDELINPEFNKKEQDFLAKRSKINLFT	212
QY	163	SRKRRTSAGTYVDBDNDNGIPDLSELEGTYVDVNNKRTLSFWISNHERKGLIYKSS	222
DB	213	QKKKREIDED--TDIDGSIPIIDLEMNGTII--QRIAYKWDQL--ASKNGITAYVN	264
QY	223	PEKMSYASDPSYDERKVTGRIDKNVSPFARHPLNAYVYUWNGMIIINSEKQSTQWT	282
DB	265	PLSHRYVGPPTIDKERANDLISNKEETNPVLAFFSVNEMEKVILSPNELTS---	320

```

QY 283 DSEFTISKNTSRHTSEVHNAEYHASFIDIGSAGFSNSNS---TVAIDHSLS 339
DB 321 ----NSVSHSSTINMSYT-----NTE-----GASVEAIGPKGISGVSVNVOHSET 363
OY 340 LAGERTAEFMG-----LNTADTARLANIRIVYNTGTAIVYVLEPTTSVLVGNKOTLATIK 395
DB 364 VAQE--WGTSQNTSOPNTASAGILANAVRYNNGVGAIVDKETTSFVL--NNDITATIT 420
OY 396 AKENQSLQILAPNNYPSKMLAPIALNADDPSPPTIMANNQCLELEKTROLRLDIDOV 455
DB 421 AKNSYALNISPGEISYKKGNGCIATISMDDFNSHPITLNNKQVDNLNNKPMLETNOJ 480
OY 456 YGNIAFYNGRVRVDTSNMSEVLPJOIQTETARIIFNGKDLNVERRIAANVSPDLE 515
DB 481 DG---YKIKDTHGNIVTSGEMNGVIOQIKAKTASIVDGE--VAEKRYAAKDEYNEED 536
OY 516 TTRPDMLTKELKIAF--GFNEPNCNLOYOGKITEFDF--NFOQTSQNIKNQIAEL-- 569
DB 537 KT-PSLTKDALKLISYDEIKTEIBGLIYKKNKPIYESSVMTYIDENAKKAVKROLNDITG 595
OY 570 ---NATNIYVLDKIKLAKMNLILNDKRFHRYNNRIAVGADSEVVEARREYINSSTEG 626
DB 596 KFDVSHLYDV---KLTPKMYVITK--LSILYDN--AESNDNSIGKMTNTNIVSGNNG 647
OY 627 -----LLNLND-----KDIRKILSGYVEIEDE-----GLKE 654
DB 648 KROYSSNPDANLTLNTDAOEKLNKNRDYYSLSYKMSSEKNTOCETITDGEIYPIITTKYV 707
OY 655 VINDRDMNL--SSLRODKTEFIDFKYNDKLPYLSNPNKYVAVYAKENTITINSE 712
DB 708 VMDKNTKRDIYAHNIKSNPISISHIKT--NDEITLWLDISI--VDVASIKPEN--LTDSE 763
OY 713 NGDT-STNGIC--KILIFSKNGYEIGZ 736
DB 764 IKOYISRYGKIKLEDDILLDKKGIIHYGE 791

RESULT 12
US-08-471-033-5
: Sequence 5, Application US/08471033
: Patent No. 5770696
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Kozlial, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalin M
: APPLICANT: Kostichka, N. Kistcy
: APPLICANT: Duck, Nicholas B
: TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESS: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NJ
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: PC DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/471,033
: FILING DATE: 09-SEP-1994
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-033-5

Query Match 20.9% Score 790; DB 1: Length 884;
Best Local Similarity 30.4%; Pred. No. 1, 18-47;
Matches 248; Conservative 130; Mismatches 292; Indels 146; Gaps 34;

4 QENRLNSESSESGILGTYFEDLNFQAPVYVTSSTGDLSPSELEN--IPSENGYFO 61
DB 42 QKNQ--QKENDRGILGTYFGKDF--SNLTMPATYRSTLIYVNCYANKLLDKQOEYQ 97
OY 62 SAWSGFTKVKKSDYETPATSDNRYVMWYDOEVNKNASNSKTYREKGRLOYKIOYO 121
DB 98 SIMIOLQSKETDPTFNLSDDEBATEIENGKIISNKGKQYVHLEKGLVPIKIEYO 157
OY 122 RENTPTKID-----FKLYWDSNKKREVISDNLOJLPELKOKS-----N 162
DB 158 SD--TFNIDSKTEKELKFKIDSGNQOQVQDELBNPEFNKESGEFLAPSKINLT 215
OY 163 SRKRSAGTPVDRNDGIDPSLEVEGTYDVKKRTPFLSPWISNHHKKGLKRYSS 222
DB 216 QKKRRIED--IDTQGSIDPLWEENGTY--QNRIVAKWDSL--ASKGYTRFVSN 267
OY 223 PERKSPADPSDEKRYTGRIKNTVSPARHPVLAAYPIYVHDENTILSKNDOSTONT 282
DB 268 PLSHVGPYTDYDEKARLDLSNAKETFNPVLAAPSVNVSMEKYILSPENLS---- 323
OY 283 DSEFTISKNTSRHTSEVHNAEYHASFIDIGSAGFSNSNS---TVAIDHSLS 339
DB 324 ----NSVSHSSTINMSYT-----NTE-----GASVEAIGPKGISGVSVNVOHSET 366
OY 340 LAGERTAEFMG-----LNTADTARLANIRIVYNTGTAIVYVLEPTTSVLVGNKOTLATIK 395
DB 367 VAQE--WGTSQNTSOPNTASAGILANAVRYNNGVGAIVDKETTSFVL--NNDITATIT 423
OY 396 AKENQSLQILAPNNYPSKMLAPIALNADDPSPPTIMANNQCLELEKTROLRLDIDOV 455
DB 424 AKNSYALNISPGEISYKKGNGCIATISMDDFNSHPITLNNKQVDNLNNKPMLETNOJ 483
OY 456 YGNIAFYNGRVRVDTSNMSEVLPJOIQTETARIIFNGKDLNVERRIAANVSPDLE 515
DB 484 DG---YKIKDTHGNIVTSGEMNGVIOQIKAKTASIVDGE--VAEKRYAAKDEYNEED 536
OY 516 TTRPDMLTKELKIAF--GFNEPNCNLOYOGKITEFDF--NFOQTSQNIKNQIAEL-- 569
DB 537 KT-PSLTKDALKLISYDEIKTEIBGLIYKKNKPIYESSVMTYIDENAKKAVKROLNDITG 595
OY 570 ---NATNIYVLDKIKLAKMNLILNDKRFHRYNNRIAVGADSEVVEARREYINSSTEG 626
DB 596 KFDVSHLYDV---KLTPKMYVITK--LSILYDN--AESNDNSIGKMTNTNIVSGNNG 647
OY 627 -----LLNLND-----KDIRKILSGYVEIEDE-----GLKE 654
DB 651 KROYSSNPDANLTLNTDAOEKLNKNRDYYSLSYKMSSEKNTOCETITDGEIYPIITTKYV 710

```

QY 655 VINDRYMLN-----ISSLRDQKTFIDFKKNDKLPYISNPNKYVAVYATKE 704
 DB 711 VINDRYMLN-----ISSLRDQKTFIDFKKNDKLPYISNPNKYVAVYATKE 760
 QY 705 NTIINPSENGDT-STNGIK---KILIFSKRGYEIGZ 736
 DB 761 N--LTDSEIKOYISRGYIKLEGGILIDKRGIGHYGE 794

RESULT 13

US-08-471-044-5
 : Sequence 5, Application us/08471044
 : Patent No. 5840868

GENERAL INFORMATION:

APPLICANT: Warren, Gregory M
 APPLICANT: Kozziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalin M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,044
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: 40,403
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8582
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 884 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-471-044-5

Query Match 20.9%; Score 790; DB 2; Length 884;
 Best Local Similarity 30.4%; Pred. No. 1, le-47;
 Matches 248; Conservative 130; Mismatches 292; Indels 146; Gaps 34;

QY 4 QENRLNSESSESGILGTFESDLNQAPAVVTSSTGDLSPSELEN--IPSENGYQ 61
 DB 42 QKMQ---QKMBRKGLGTFYFKGKDF--SWLTFATPSTLIYDQVTANKLIDKQDEO 97
 QY 62 SAWSGCFIKVKSDEYTFATSDNHNWYMWVDDQEVINKASNSNKILFELGRLYQIKYO 121
 DB 98 SIWVIGLQKQKQFPTFNLSEDOALIEINKIISNKGKQVHLEKGLVYIKIEYO 157
 QY 122 RNPFTKGLD-----FLYWTDSQKKREYISDNLOLPFLKOKS-----N 162
 DB 158 SD--TKFNIDSKTFKELAFKISQNPQQVQODELNPFPFNKESQELAKPSKINLET 215
 QY 163 SRKRSTSAQPYVPDNDGIDPSLEEGYDYVKNKREFLSPJISNHEKKGLTYKTS 222
 DB 216 QKMRREID--TPTDQSLPDLMEGTYT---QNRVAVWDSL-ASKGYTFVSN 267
 QY 223 PKKSTASDYSDFEKVYTGRIKDNVSPKARHPLVAFIVYVDMENILSNREQSTQWT 282
 DB 268 FLESHVGDYDYDEKARLDLSNAKEFNPLVAFPSVNSMEKVLISNENLS---- 323
 QY 283 DSERTISKNTSTSRHTSEVHGNAEYHASFEDJGGSVSGFSNSNS--TVAIDHSLS 339
 DB 324 ---NSVEHSSSTNMSYT-----NTE-----GASVEAGIGVMSISPGSVNYOHSET 366
 QY 340 LAGEKTAETMG---LNTADPAPLANANIRYVNTGAPVYVLTSLVLCNKQTLATIK 395
 DB 367 VAQE--WGTSTGNTSQFWTASAGYLANAVYNNVGTGATDYKFTTSPVL--NNDTIATIT 423
 QY 396 AKENQSLQIAPNNYPSKNIAPIALNADDFSPPTMANYNOFLEKTKQLRPTDOV 455
 DB 424 AKSNSTALNISPESYPRKGNGIATISMDNPSHTTLNKKQVNLNLNKPMLLETNGT 483
 QY 456 YENIATYFENGSRVYDTSQWSSEVLPQOETTAIIFNGKOLNVERIAAVPSDPLE 515
 DB 484 DG---YKIDTHGNIVTGGEMNVOIQKATASIIYDGE--RAEKRVAAKYNPBD 539
 QY 516 TTKPMTKAKIKAF--GFEPNGILOYGKDTEFD--NFDQTSQINKNQLAEL-- 569
 DB 540 KT-PELTKALMLSTPDEIKELBGLTYKNKPILESSVMTYIDENYAEVYKQDLDTG 598
 QY 570 ---NATNIVYLDKIKLAKANMLIDKRFHYDRNNINAVGADSVYKAEHVEVINSSTG 626
 DB 599 KFDVSHLYV---KLTPKANVYIK-LSLIDN---AESDMSIGKMTNINISGGNNG 650
 QY 627 -----LLNID-----KDIRKILSGYVEIETE-----GLKE 654
 DB 651 KKQYSSNPNPANLTNTDQKLNKNDYISLYKSEKNTQCELTIDGTYPTTKTYN 710
 QY 655 VINDRYMLN-----ISSLRDQKTFIDFKKNDKLPYISNPNKYVAVYATKE 704
 DB 711 VKDYKRLDILAHNIKSNPISSLH-----IKTNDETLFWMDISI--TDAVASIKPE 760
 QY 705 NTIINPSENGDT-STNGIK---KILIFSKRGYEIGZ 736
 DB 761 N--LTDSEIKOYISRGYIKLEGGILIDKRGIGHYGE 794

RESULT 14

US-08-463-483A-5
 : Sequence 5, Application us/08463483A
 : Patent No. 5840870

GENERAL INFORMATION:

APPLICANT: Warren, Gregory M
 APPLICANT: Kozziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalin M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE: 05-SEP-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CITP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8615
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-483A-5

Query Match 20.9%; Score 790; DB 2; Length 884;
Best Local Similarity 30.4%; Pred. No. 1.1e-47;
Matches 248; Conservative 130; Mismatches 292; Indels 146; Gaps 34;

QY 4 QENRLNSESSESSGILGYFSDLNFOAPAVVTSSTGDLSPSELEN--IPSENQYEQ 61
DB 42 QKNO---QKMDRKGLGTYFFKGNF-SNLMPAFPRDSTLIYDOTANKLDKQOEYO 97
QY 62 SAHSGFIVKRSDEYFATSADNHVYVADQEVINKASNSKIRLEKGRLYQIKIOYO 121
DB 98 SIRHIGLDSKETGPTFNLSDEQALIEINGKIIISNKGKQVHLKGLVPIKIEYO 157
QY 122 RENTPEKGLD-----FKLWTSQNKREKVISNOLPELKQKSS-----N 162
DB 158 SD--TKRNIDSKTEKELKFKIDSONQPOVQOGLNPEFNKESQEFLLAKPSKINLFT 215
QY 163 SRRKRSAGPVPRNDGIPDSLEYEYTYDVKNKRTLSLSPWISNHEKGGITKSS 222
DB 216 QMKRREIDED--TPTDSDSPDLMEKSGITL---QNRLAYKMDDSL-AASKGTFKYSN 267
QY 223 PEKSTASDPSYDPEKVTGRIDKNVSEARHPLVANPIYHVMEMIIISNENQOSTQNT 282
DB 268 PLESHTVGDYTDYEKAARDLDSNAKEFNPPLVAFPSVNMENKYLISNENMS----- 323
QY 283 DSEHTTSKNTSTSRHTSEVHGNAEYHASFDDIGSVSAGFSNNSNS---TVALDSLS 339
DB 324 ---NSVESHSSTNMGSYT---NTE-----GASVEAGISPKGISGVSQVNTQSHET 366
QY 340 LAGESTAPRTEWG---LNTAOTARLNANINIRYVGTSTAPIYVNLPTTSILVGNKQTLATIK 395
DB 367 VAOE--WQTSIGNTSQFNTASAGYILNANRYNNGVGAIDYVKETPSFVL--NDNTIATIT 423

QY 396 AKENOLSQLAPNNYTPSSKNLAPALNAODESSPTIMANYO-TLEKTKQLRDLTDYV 455
DB 424 AKSNSTALNISPGESEYPPKRGONGAIYTSMDFNBSHPTLNKKQVNLNKKMLETNT 483
QY 456 YGNIAVYFENGVRVDTGSNNSEYLPQIOETTARIITNGKOLNLEKRIAVNSDPLE 515
DB 484 DG---YKIKIDTGHINITYTGEGWNGVIOQIKAKASTIYDGE-SVAKRYAKQEDNEP 539
QY 516 TPKPDMTLKALKIAF--GFNPNGNLOYGKDIETFE--NEFQJISONIKNOALF-- 569
DB 540 KT-PSLTLDKDLKLSYDEIKETEBELLYKKNRPYESSVMYTLIENTAKFVKOLANDITG 598
QY 570 ---NANITYYDKTKINAKMNLIDRKRHYDRNNIYVAGADESVYEAHREYINSSEG 626
DB 599 KFDVSHLDV---KLTPKMYTITK-LSLIYDN--AESNDNSGKMTMTNIVSGNGG 650
QY 627 -----LILNID-----KDIRKILSGYIEIDTE-----GLKE 654
DB 651 KKOYSNNNDALTLNTDQAEKLNKRPVYISLYMKSEKNTQCERTIIDGEIYPTTKYVN 710
QY 655 VINDRDMNL-----ISLRDDGKTFIDFKKYNDRKLPVYISNENRYVAVYKRE 704
DB 711 VNKNTRKLDITLAIHINKSNPLSLH-----IKTNDITLWDDISI-TVVASIKPE 760
QY 705 NTIINSENGDT-STNGIK---KILFESKGEIGZ 736
DB 761 N-LTDEIKQIYSRIGIKLEDGILDKKGIHCE 794

RESULT 15
US-08-471-046A-5
Sequence 5, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas J
TITLE OR INVENTION: Method For Isolating Vegetative Insecticidal
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 5866326artis Corporation
STREET: 3054 Cornwalls Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057

Fri Oct 4 14:31:03 2002

us-09-848-909-21.ra1

Page 12

FILED DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIY8 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-046A-5

Query Match 20.9%; Score 790; DB 2; Length 884;
Best Local Similarity 30.4%; Pred. No. 1,1e-47;
Matches 248; Conservative 130; Mismatches 292; Indels 146; Gaps 34;

OY 4 QENRLNNESSOGILGYVPSDLNFOAMVYTSSTDLSIPSELEN--IPSENQYFO 61
DB 42 QKNQ---QKENDKGLGYFKGDF-SNLTMPAPTRSTLITDQOTANKLDRKQOEYO 97
OY 62 SAIVSGPIKVKKDEYTPATSDNHVYMWVDOEVINKASNSKIRLEKGRLOYKIOYO 121
DB 98 SIRMIGLIQSKETGDFMLSEDOALIEINGKIISNCKEKOVHLEKGLVPIKIEYO 157
OY 122 RENPTEKGLD-----FKIYTDSONKKEVISSDNLOLPELKOKS-----N 162
DB 158 SD--TKFNIDSKTFKELKLEKIDSONOPQVQODELRNPFNKKESQEFPLAPSKINLFT 215
OY 163 SRKSTAGPTVPDRNDGIPDSLEVEGYTVYVKNKRTPLSPWISNHEKGLTKYKS 222
DB 216 QKKAREIDED---TDVDSGSLPDLMEWGTIT---QNRIVAKMDSL-ASKGYTKFVSN 267
OY 223 PERKSTASDPYSDEKYTRIDKNSPBAHPVLAAYPIVAVDMENITLSKEDOSTONT 282
DB 268 PLESHYVGGPTDYDEKAAARDLSNAKETNPVLAAPSVANSEKYLSPENLS----- 323
OY 283 DSETRISKNTSTSRKTSVHGNAEVHASFDDIGGSVAGFSNSNS---TYAIDHLS 339
DB 324 ----NSVESHSTWMSYT-----NTE-----GASVPAIGPKGISRGVSINQIHSET 366
OY 340 IACERTMAETMG---LNFADTARLANIRVYNTGAPLYNVLPPTSVLVGNQTLATIK 395
DB 367 VAQE--NGTSTGNTSOPNTASAGYLANAVRYNNVGTGAIYDKPTTSFVL--NNDTLATIT 423
OY 396 AKENQLSQILANNVYPSKMLAPIALNADDSSTPTIMVNOPELEKTKQLRLDTDOV 455
DB 424 AKSNSTALMISGSEYPRKGGONGIATTSMDPNSHPYTLNKKOYDNLNKKPMLMETNOT 483
OY 456 YGNLATNPNENGVRVVDGNSWSEVLPJOETARLIFNGKDLNIVERRIAANVSDEPLE 515
DB 484 DG---YKIKDWHGNLVITGGWNGVIOQIKAKTASTIIVDGE--RVAEKRYAARDYENPED 539
OY 516 TTKEDMTLKPAKIAF--GENEPNGLOYOGDITEEDF--NPDQOTSQNIKNQLAEL-- 569
DB 540 KT-PSILKNAKLSTYDEIKETEGLLYKKNKPIYESSVATYLDENTAKETKOLNDITG 598
OY 570 ---NATNIVYADKIKLAKKNLITDRKFRHIDRNNTIAVGADESVVKAHREVINSSTEG 626
DB 599 KKVDSHLYV---KLTPKNVYTIK-LSILDN--AESDNSIGKWTNTNIVSGNGG 650
OY 627 -----LLMID-----KDIRKILSGYVLEIDTE-----GLKE 654
DB 651 KKOYSSNPNPAALTLMTDQEKLNKNDYISLYMKSEKNTQCEITIDGELIYPTTKTYN 710
OY 655 VINDRTDMLN-----ISSLRQDKFTIDFKKYNDKPLPYISNPNVYVYAVATKE 704
DB 711 VVKDNYKRLDILAHNIKSNPISLH-----IKTNDLITLPMDDIST--TVVASIKPE 760

OY 705 NTLINSENGDT-STNGIK---KILIPSKGYEIEGZ 736
DB 761 N--LTDSEIKOYISRGVILKEDGILIDKKGSLIRHGE 794

Search completed: October 4, 2002, 14:34:45
Job time: 30 sec

Fri Oct 4 14:31:03 2002

us-09-848-909-21.ra1

OY 370 GTATINVLPTTSLVIG-----NOTA-----TIAKENG-LSQILAPN-- 409
 Db 541 GEA-----SLVAERLDINADIKTINNGTLAGLANTITKALENRDVALILAHOMIN 591
 OY 410 -----Y-----PSKNAPALNADDPSS-----TPTMYNNOFLELEKTOF 448
 Db 532 FTYNGSHYVNRKDIYSKDAIVTFSSNMDFTSNSSKDYADQNNILVAVNNENFTGSGEII 651
 OY 449 RLDPDQYTGNI---ATYFEN-GRVNDGSGNSEVYLPDQCTARILFIIFGKQMLYER- 503
 Db 652 -----LHGAVLNAKNGFTSGNGLTYKHEINISINIST---NAGNLTGKMLFVHSNT 702
 OY 504 -----RIAAVNSPPLTEKRPMTL-KEAKIAPG---FNPNGNIOYOGKIDIEF 550
 Db 703 TVKNDGLVSIENLISKTDFPTNGSTLGLLEALKLASGNTNMSGSLA-SNKSLDIY 761
 OY 551 DENF-DOOTSQNIKQIALMELNATNIYYIDK--IKLMAKNLILDRKFRHDDNNIAYVA 607
 Db 762 GNNITNGTLETSKS---LNTINNTFTINNATIKSYGVALNTISG-NTINDSGVMSH 816
 OY 608 D-ESVVEAHREYINSST---EGLL---NIDKIRKILSGIYELEDTEGLAKYIND 658
 Db 817 DLINITSQAN--IINKNLACGGQGLNLTAKGITINDSNSTALAVLSNNDIN-----LWA 869
 OY 653 RYDMLNINISL-RODKTEPIDRKYNDKLPLYSINPNKYVAVYKENTLINPSENGDTS 717
 Db 870 NNRKYVINIGEIYSQAGNISVEAKLILHDVXL-----SGNITTTTK-----SGNATVK 915
 OY 718 TNGI 721
 Db 916 TNSI 919
 RESULT 10
 R90598
 membrane nuclease, lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: E90598
 R:Chamblard J., Heallig, R.; Pettis, S.; Barbe, V.; Samson, D.; Gallisson, F.; Moszer, I.;
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul-
 A:Reference: A99512; M01D:21267165; PMID:11355084
 A:Accession: E90598
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1123 <GCR>
 A:Cross-references: GB:AL445566; PID:gl4090108; PIDN:CA613866.1; GSPDB:GM00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MRP0_6930
 A:Genetic code: SGC3

```

Query Match Similarity      4.88; Score 182.5; Db 2; Length 1125;
Best local similarity 19.48; Pred. No. 0.16;
Matches 167; Conservative 154; Mismatches 341; Indels 197; Gaps 39;

OY      3 KQENLLINSESSGCGLLGYFFSTINFGQAPMYVNSSTGSLSPSELEINPSEMYFOS 62
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      120 EMDNSVNNKKNSSK-----NDEN----LTLTKVGLGMNV-LNQGSLPFRKLNLISK 168

OY      63 AIWS-----GFIKV-----KSPDETFATSDNHYTMWVDDE-----VI 97
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      169 VILHKDLVGLLEITINNGYKTTINELMKSSDCKMNYISLKKGTGGSGGDEHVGII 228

OY      98 NKASNSKTRLE-----KGLRLOIKIÖYORENPFEKL----- 130
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      229 YK-----ENLLTLESPFDKREKSGKRYETKNTM---DDPFRKÖKÖITDIYVPEYGFSTKNGI 282

OY      131 --DFLYLV--TDSÖNKR--VISOONLÖLPPELÖKÖSNSRRKRTSAGPVPYDRNDÖI 183
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      283 KIDFIVVPMHSDDPVGVEKEEISIAKQVSGÖGCKRYVADALKTREYMETFDSI-----DÖV 337

```

```

0Y 184 PDSVNGYV-----VDYKNTKTELSF-W--ISNTHK---K 214
Db 338 NNELEFPGDNIKNIKONEKKAPEFLLOSGKSLIDVENVNTSLVORGEVANHDXKLEK 397
OY 215 GLTKYKSSP--EKWSTADSPYDSEKNGYGRIDK-VNPEARHP-----LVAAFYI 261
Db 398 GDUKXENSGYDLMKYPEDDNLNKEEKKEKRYEBSKAKAKDKTGEYSYVLAIHSDHYI 457
OY 262 VHDVEMNILLKSNEDOSTOFTDSETRTKSTSTSRKTHSEVHA-EVHASFDDIGSVS 321
Db 458 VYTDI--TILQKND--ONKSENK--DENSDSKOHLDKRTINEONONTODDKKIS 509
OY 322 AGFSNNSSTVAIDHSLSLAGERTWAGTLMFADTALNANI-KVNTGTPAYINLEPT 381
Db 510 DKSQNSNTWNEKDOKLDDCODE--SKNNAIKISQNDOKOSINLSSKNDPTOKRESSPOI 566
OY 382 SLVLKKNQNTATIKAKENQLSQILAPINNYPSKN-----LAFIALNADDSSTPTYN 435
Db 567 NNLENNQDEISHNNGENDSKBEONTNSRQTKNLSBEOKONLHPKPNSSNSETNEN 626
OY 436 YNOLFLEKTKOJRLDPTOYGNATJATNEGNORVAPDGSNMSW-POIOTETARILFNG 495
Db 627 EYONNENSTIKKEIDTSAKTODSTNSNLKN-----EKTINQEKRTIESNNS 675
OY 496 KDLNIVERLIAVNPSPLEETFKDPMULKEA-LKTAISEPNPCH-OYOGODITEPFPNR 554
Db 676 NSTNQOE-----NSS--TKKEIKESSEVNNNSNTKOGEMINDKKEIKESPSNV 726
OY 555 DQOTSQNIKN-LAELNATNYIYLDKIKLNKAKN-----ILLJCKRPHYDNNIAANGD 608
Db 727 NNSNSTNQNONETPENNESONNVITIGNPNOSLQMONLIDSKAKYKIGIWNINESGVS 786
OY 609 ESYVEAREVJINSTEGLL-----LNTDKDIRKILSGIYEIEHTEGLEKEVINDRYDMA 664
Db 787 SAKSAVAAYAKIDHNKLDLYGIGIGIVAEHELTKI-----VEBNKLKSKOSDKW--VO 837
OY 665 ISSLRDDKPT-IDFKYNDKRLPYIINPYKANYAVYTKENT--IINPSENDITUNG 720
Db 838 VISEKKEGEPVNLARYIG-VIYKEN--KNIIESKRONTPKCHLEYENOPNNSSFT 892
OY 721 IKKI-----LFSKKG 731
Db 893 SEKVSYVNPFGIKESTKKG 911

```

RESULT 11
T28677
rhoptry protein - Plasmodium yoelii
C.Species: Plasmodium yoelii
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #tax_change 01-Dec-2000
C.Accession: T28677; C45521
R.keen, J.; Stinpa, K.; Brown, K.; Holder, A.
Mo. Biochem. Parasitol. 65, 171-177, 1994
A.Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii
A.Reference number: Z20508; PMID:95021522
A.Accession: T28677
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-2269 <KEE>
A.Cross-references: EMBL:U27838; NID:9457145; FID:9457146; PIDN:AAA21304.1
R.keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mo. Biochem. Parasitol. 42, 241-246, 1990
A.Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A.Reference number: A45521; PMID:91101660
A.Accession: C45521
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 2131-2269 <KE2>
A.Cross-references: GB:M54283

Query Match	4.88; Score 181; DB 2; Length 2269;
-------------	-------------------------------------

Best Local Similarity 18.7%; Pred. No. 0.54; Indels 214; Gaps 36.
Matches 158; Conservative 140; Mismatches 311.

```

QY 1 ENYGEURL-----LNESESSGGLGYYFSDLPFOAPMVVTSSTGDLSPSELENIPSE 56
DB 485 OTLEKKTETESLANHANNNELI-KYSDKLANGLINEENGLNOTETEKTFNDIEK 543
QY 57 NOYFOAHSOFIVKKSDEYFATSANHYTMVWD-----OEVIKANSNKRILEK 110
DB 544 NIHNEIKETIKIHS-IYINSEFEIREIGINISLTKVEKEVENTINIKRK- EK 601
QY 111 GRLOYKIOYORENTEKGLDFLYVYOSK-----KEVISDMLQ----- 152
DB 602 LKHTDF-----SDPKGKIKYTKIKKINDIMAVSQIDIOHINOLDIOKSES 652
QY 153 -LPELKOKSSNRKKTSTAGPTVPDRNDGIPDSLEVEGYVDYKPKRFLSPMISNH 211
DB 653 YVSEMKOINKLEKVSNT-----EISND-----NVEG-----KKKQOIT---VTKID 692
QY 212 ERKG-----LTKYKSPKSTASDPYSDPEKVTG-----RDKRVSPPEARHPL 255
DB 693 KKKNIYEIKLISEIKEDNTSEKAYDINISYONGLNIFLEQIDEE-KKKAKNTI 751
QY 256 VAAAPYIVHDMENILSKNEOSTONTDETRITSKNTSTRTHTSEVHGAEVH---A 311
DB 752 KSEMEAYD-DLDNIRKKSQETETEMDKMDIKEMELAKISHDDCKCHDKSKKHKENS 810
QY 312 SFEDIGSVAGFSNSNSTVAIDHSLAGERTVAETMGNTADATLANINITYVGTG 371
DB 811 DLYDKSSKIIQDFSR-ESDINDIRKILQ-----KNSSESQNHNDIOCLNEY 857
QY 372 APTVAVLPETSLVIGKNOTLATIRAKENOLSOILAPNNYPSK-----HLAP 418
DB 858 ANIYHIL-----KLNKIKKIDIKREYTSIELEKKNKINIDELANSER 899
QY 419 IALNODFSTPTIYANNOFLE-----LEKTKOLRIQD-----QYGN 458
DB 900 VIKKIEGSLSEKCRKINSITLDOKDIDECIKININILKKNLNEININIHKKAEIK 959
QY 459 IATYNEGRVAVDGSNMSEVLPJOIETTARIIFNGKDLVERRIAVANPSDPLETK 518
DB 960 IVLSNHN-----IMADNKSQYILIEIKKNGT---NDHYNIKELK-SHKDSNRYK-TE 1010
QY 519 PDMTLEAKLAFNPNPONGNLOYGKD-----ITFEFDPFOQTSOINIKOLAE 568
DB 1011 ADONKAIOR-----NKELEFYKREYVLYNKYVAVELKKNKF-KTKNDSKOIIRE 1061
QY 569 IMAITTYVULKIKLANKN-----ILIRDRFHYDRNNIAGADESVYKAEAREVIN- 621
DB 1062 IKDHANCTLESCKSEKKNKINEKIHIEDEVANDNKNKAITSIKYVEPEPKTKIKI 1121
QY 622 ---STEGLLNIDKIDIKILSGYVEIJDY-----EGKFEVINDRYDMILNISLR 669
DB 1122 NEIFKSDCLKETNDEKQISLSDIOETKRLNENCKOLKTLLEL-----LESIK 1172
QY 670 ODGKTFIDEKTYNDKPLIYISN-----PNKVVAVYVREKNTIIPSENGDSTNG 720
DB 1173 KOKKNIDOKREIDEVNSKINIENTVNOHKNKNEIIGI---VEKINIAKTKNQISSTRE 1230
QY 721 IKK 723
DB 1231 LIK 1233

```

RESULT 12
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C:Accession: S05603; S04850
R:Myler, P.J.
Submitted to the EMBL Data Library, April 1989

A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:99896; PDB:1CNA33163.1; PID:99897
R:Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from
A:Reference number: S04850; MUID:89345116
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL>
A:Cross-references: EMBL:X15063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-1639/Product: major merozoite surface antigen #status predicted <Mat>

Query Match 4.8%; Score 179.5; DB 2; Length 1639;
Best Local Similarity 20.4%; Pred. No. 0.39;
Matches 171; Conservative 111; Mismatches 272; Indels 285; Gaps 41;

```

QY 5 ENRL-LNESESSGGLGYYFSDLPFOAPMVVTSSTGDLSPSELENIPSENOYFOA 63
DB 927 ENILSLGKKNKIYDELIGKSE-NF-----YKIKDSOTFVE 965
QY 64 TWSEFTYKKSDEYFATSADNHVWVDOOEVIKANSN-----NKIRLEKGRLY 114
DB 966 SFTEFVSKRAD-----INSLNDSKRRKLEEDINKTKTLQISFDLYNKTKLDERLD 1020
QY 115 -----QIKYOARENTEKGLDKIYWDSONKEVISSDNLOPELKOKSSNSRKRS 168
DB 1021 KKKTVGYKKMOIKKILTKLEDESKL--NSLNPKHYL--QNFVFPKKKEAEIETEN 1076
QY 169 TSAGPTVPDRNDGIPDSLEVEGYVDYKPKRFLSPMISNHEKKGGLKY---KSSP- 223
DB 1077 T-----LENTKILKHY-----KGLVYKNGESSPLK 1103
QY 224 ---EKWSTASDPYSDE-----KVGRIDKAVSPEAR-----HFLVAAPIYVHD 265
DB 1104 TISESIOETEDNTASLENKFKYLSKLEKLDNINSEKKSLSYLSQDHLHLA---E 1156
QY 266 MENILSKNEOSTONTDETRITSKNTSTRTHTSEVHGAEVHASFEDIGSVAGFS 325
DB 1157 LKEVIKKNK---YGNSPSENNY-----DVNNALESIKKFLPEGTDAVAVS 1200
QY 326 NSNSSTVAIDHSLAGERTVAETMGNTADATLANIRVYNGTAPTYNLPPT----- 381
DB 1201 EGSODTLPOQPKKPASTVGAES---NITTSQ--NVDEVDVYIVPILGESEDDYDL 1256
QY 382 -SLVIGKNOTLATIRAKENOLSOILAPNNYPSKNAPIALADODFSSPTIMYNOL 440
DB 1257 GOVYGEAVTSYI---DNILSKI---ENEYEVLYLKPLA-----GYR 1294
QY 441 ELEKTKQLADTDVYGNINATYVFNGRV---RVDTGSNMSEVLPJOIETTARIIFNGK 497
DB 1295 SLRK-----QLENNVATEFVYVVKDILNSRKNREKFNKVLH-----SD 1332
QY 498 L---NLYERRIAVANPSDPLETKRPMTLKALAFNPNPONGNLOYGKDITEFDN 553
DB 1333 LIPKDISSNRYVADPYKFLNKEKDKLSYVI-----KSDITFDIN 1377
QY 554 FDOQT-----SONIKMOLAE-----NATNLY-TVLDKI----- 581
DB 1378 FANDVLYTKILSEKKSDDSIKKIINDGEGENKTLPLNINIEFLKTYVNDKTDIFVI 1437
QY 582 -----KLNKAKNTI-----LIRDRFHYDRNNIAGADESVYKAEAREVINS 622
DB 1438 HLEKAVLNATYTESNVEYKIKELNLYKTIQDLDKFNKNNFVGIADLSTDYNNINILTK 1497
QY 623 ---STEGLLNIDKDI-KRILSGYI---VEIETDEKLEEVINDRYDMILNISLRDQKTFI 676

```

Db 1498 FLATGAVENILAKTVLSNLDGMLNLSOQCKKCKPONGSGCRHLDERECKCLL 1557
 QY 677 DFKYNDKPLPLISNPNYKAVYAVKENTININSENG-----PFTNGIKKI 724
 Db 1558 NYKQGDGKC---VENPNTCN-----EN-----MGCGADAKACREDSGSGNG-KKI 1599

RESULT 13
 A:Accession: AB6827
 C:Species: Lactococcus lactis subsp. lactis (strain IL1403)
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 R:Botulin, A.; Mincer, P.; Manger, S.; Tallon, O.; Malarme, K.; Weltschbach, J.; Ehrlich
 Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s.s.
 A:Reference number: AB6825; MUID:21235186; PMID:11337471
 A:Accession: AB6827
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1072 <STO>
 A:Cross-references: GB:AE005176; PID:912724625; PIDN:AA05715.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: yqfg

Query Match 4.7%; Score 178.5; DB 2; Length 1072;
 Best Local Similarity 19.4%; Pred. No. 0.24; Indels 179; Gaps 30;
 Matches 160; Conservative 142; Mismatches 342

2 VKQENRLNESSSSOGILGYPSDLAFQAPVYTSSTIGLSTIPSEELN----- 52
 300 VEQYDVASSSETPQDANSASLYPISEASVYDNTLMSLSLDSISSGOTENSGASSTA 359
 QY 53 ---IPSNQYFOSA---IMSGFTKVKDEYFATSDNHYVMWDOEYVKNASNNKI 106
 360 EIVSDSESNLSSSQINSNSNSENKOSLSGSSHESEHNSNNETNNSEI 419
 QY 107 R-----LEKRLVQIKIYQORENTEKGLDFL---YMDSONKKEVYISDNQLPELK 157
 420 TILPSPNTESSNSVSDQSSSEASTNSNISLSNISTSYSEANNSDFENAEVA 479
 QY 158 OKSSNRKRKSTSNAGPTVPDRNDGT---PSLEVEYTVQVKNKRFPLSPWISNHEK 213
 480 NNSLASVNNSSSVLSTSTADNGLINOSGDLTKD---SEISTSGAPLS---SNQTS 534
 QY 214 KGLTKYKS---SPEKMS-----TADPYADEYKTVGRIDKVNPPARPIVAAVPI 261
 535 EASTNSNSSLSPSNISLSTVLESTISSNPFENAEVANNSLASVNNSSSVLSTSTA 594
 QY 262 VHYDM-----ENT-----ILSKNDQSTQNDSEFTPTISKNTSTRTSTSEVH 304
 595 DNEIINQFSGDMLTKRDSSETISGAFLSNQTSSEASMSNSINSPLSLTSLTSSSA 654
 QY 305 GNALVHASFPGDIGSVAGSNSNSTVAIDHSLT---AGEETAELMGLMTADTARLNA 362
 655 TNO-----SNSRATVYDNNST---HSSNLTNGSDSDSDSDSDSDSNNL-- 699
 QY 363 NIRVNTGTAPLYNVLTPTSLVGRKNOTLATIKKENOJSQ---ILAPNNYPSKNLAP 418
 700 -----SSSNLETPNOTISSKPEVNNISEPPKVVSSSVENSTTHE 742
 QY 419 IALNADDPSTPTITMANTNOLEKTKQLRLDQVYGYATYV---FENGVRVDTGS 475
 743 MESTNRSISSTISSTSSQOKESGN---LLNTTEGNNPITFNNSSSEMSASTLT-- 797
 QY 476 NMSEVLPOLOETIATITP-----NGKDL-----NIVERRIAVNSDPLETTPKMT 522
 798 STSNSSSTGCLYISNEARQNSGETSHSLPSNSNNENNVSSISQOALLESKSTN 857
 QY 523 LK-EALKTAFQENP-NGNLQYQCKDTEPDFNFOQTSNIRKNOALBETNNTIYVLK 580

Db 858 KXSSLSLINSTSHPNEDNOSNSD-----EVKSNNNVESILGOLNSISKTHMS 909
 QY 581 I---KLNKNNMILTRDKRPHYDRNNIAVAGDSVYKAEHREVINSTEGILLINDIDIR 637
 Db 910 LISOAKISVLYTLPSKSKVTNENKNSNVSSEKILTKPKN-DKSONLQITALLDSFNK 968
 QY 638 ILISGIVLEDETELEKEVIN-----DHYDMLNLSLROD-TRFIDFKYNDKPL 688
 Db 969 -----EVEFEDSKTVPDKVLDNEDGNSONNKSTYADKMKVFPKRSEFNKLT-- 1019
 QY 669 ISNPKNVAVYAVKENTININSENGDTSTNGIKTILPSSKC 731
 Db 1020 -----DSDNMLTKTVTLKHK 1036

RESULT 14
 A:Accession: MYPV_7030 [imported] - Mycoplasma pulmonis (strain VAB CTIP)
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 R:Chandaud, I.; Helli, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma
 A:Reference number: A99512; MUID:21267165; PMID:11353084
 A:Accession: G90599
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1752 <KUR>
 A:Cross-references: GB:AL445566; PID:914090118; PIDN:CA013876.1; GSPDB:GN00153
 A:Experimental source: strain VAB CTIP
 C:Genetics: MYPV_7030
 A:Gene: yfpg

Query Match 4.7%; Score 178; DB 2; Length 752;
 Best Local Similarity 19.4%; Pred. No. 0.15; Indels 262; Gaps 41;
 Matches 160; Conservative 125; Mismatches 277

41 GDLSPSELENIPESENOYFOSAIMSGFIKYSDEYF-----ATSAOHHTVAMD 92
 20 GTISGISYGISOLPNEASL-----YKRNLMELNKNSTYLNKSIETILNOMD 69
 QY 93 DOEYINKASN-----SNKIRLEKGR-LYQT---KLOYRENFTEKGLFWTDSQ 140
 70 EKVYNSISANDFPNKFVPSKTPLENGEKITTVSLGHOYFOLVYNSQKITSKSVKITSK 129
 QY 141 NKEVYISDNMLPELKOSN---SRKRSISAGPTVPDRND---GIPSLVEGYTV 194
 130 ISKDVAMD-----KORLNDPKNKLVNFKSSASEKQSDYVASFNDKSLKILYL 182
 QY 195 DVKN-----KRTPLSPWISNHEK---GLTVKSSPEKMTASDYP 233
 183 DKNVEKNISENTEFVELKTENNAFVGGISNDIGTVLEATVYKKNNSK-----Y 235
 QY 234 SPEKVTGRIDKNVSEARHPLVAAVYTVHDMENILISKNEDQSTQNDSEFTPTISKMT 293
 236 ONIKI---QITNNE-----KRFQISDESILKELSNVAV-- 266
 QY 294 STSRHTSEVGNALVHASFPGDIGSVAGSNSNSTVAIDHSLTSLAGERTMAETGLN 353
 267 -----TEBQKSKSEMDASFS-GMSLSKFSKLNKND---DEKIKLES---YFVDLK 312
 QY 354 TADTARLNANIRVYNGTAPLYNVLTPTSLVGRKNOTLATIKAK-----ENLQIL 405
 313 TNEKTLAKYKHDK-----FKLEIYEQKSGVTLVYLGWVAVNGESIPNHLISLF 366
 QY 406 APNNYPSKNLAPALNADDPSTPTITMANTNOLEKTKQLRLDQVYGYATYVFE 465
 367 RPS-----SKESTIITIGKEELSKRELKINSIVKVELKMTSODAD-KILASTISKQFD 421
 QY 466 ---NGRVRVDTGSN-----WSEV-LPOLOETIATITP-----RIIFNGRLNLVERRIA 506

```

Db 422 OSQANKRITIGKRAMTIDAKLAKYITELISEVNIINDTGLIKLYLFLKIDDDSTIRKTT 481
OY 507 AV-NPSDPLETTKPD-----MTLKEALKIAGFENGNGMLQYOKODITPEDFPDQOTS 559
OY 482 LVLSGFKKVATKPEDEKMDVEIKKSLMOIT-----YDDP----- 517
OY 560 QIKKQIAELNATNIY--TVLDR-----IKIAKMLILRDKREHYDR 600
Db 518 -----LNTINLYKNTITLSEKLNKRIRQSIQEDFDIKILNTNREKEDWRLDFO 568
OY 601 NNIAVGADESIVKAREVINSSTEGILLNIDKDIRKILSGYIYEIPEDEG-LKEVINDR 659
OY 569 NEKAI-----EYKNNQSSSVNFKYKVKV---VHSDKSSGKIEDIYEDK 613
Db 660 YDMLNITSLRQDG-KYFIDP-----KYINDKL-----LYISNPNKYVNYA 700
OY 614 VFSKEVKKNGNAARQDPSFLKDSQRLKPLSQITTNELKDQISYVDEPKLIDITSSN 673
OY 701 VTEN-----TINPSE--NGDTSTNGIKKIL-----FSK 729
OY 674 INRDNPFYKIEPKDKKAD-DKNGSIKIMVSLYKNTKFSK 716

```

RESULT 15

```

Jc6009
surface-located membrane protein Imp3 precursor - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: Jc6009
R:Adeloid, S.A.; Jensen, L.T.; Brock, B.; Birkeind, S.; Christensen, G.
J. Bacteriol. 178, 2775-2784, 1996
A:Title: Analysis of 0.5 kb chromosome pair repeats in the Mycoplasma hominis Imp gene system
A:Reference number: Jc6009; MIMD:96213016
A:Accession: Jc6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CA64858.1; PID:g1197336
C:Genetics:
A:Gene: Imp3
A:Genetic code: SGC3
C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homology
C:Keywords: duplication; membrane protein
F:1-74/Domain: signal sequence #status predicted <SIG>
F:75-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>
F:957-992/Domain: tetratricopeptide repeat homology <TP1>
F:993-1026/Domain: tetratricopeptide repeat homology <TP2>
F:1029-1120/Domain: tetratricopeptide repeat homology <TP3>
F:1124-1190/Domain: tetratricopeptide repeat homology <TP4>

```

Query Match 4.7%; Score 178; DB 1; Length 1302;
 Best Local Similarity 17.9%; Pred. No. 0.33;
 Matches 137; Conservative 141; Mismatches 287; Indels 200; Gaps 27;

```

OY 1 EYKQBNLLNESSESSQGLGYFDLNFQAPMVVTSSTYGDLSIPSSSELENIPSENOYF 60
Db 476 QYDENKNSITQELNMLIDKANTLLPOLINDNSEIYKAKEINAEITNANKRANVONDNASH 535
OY 61 QSAIWSGFIYKKSDEYFATSDNHNVTWVDDEVINKASNSNRIEKKR-LYQIKI 118
Db 536 QSA-----KSLDDKYTKIQONOLTEFNKDKDAKFEKELEQTRKDIDNFLT 579
OY 119 QYQENPTKEGL-----DFKLYTDSQNKKEVSISSDNILPELKOKSSNSR--KRS 168
Db 580 DQYKKNPQVATLVKDLNKAQDKKSVTKSSNKEIIAND---ELKQALDKAVAKQI 635
OY 169 TSAGPTVDDNDGIPDLSLEVGGTV---DYKNRKFLSPWISNIEKKGL----- 216
Db 636 DEKANSIKQILOSITNANQNLNKLIVSDKIDQAKTEISOIQAQSEELNANPFSMQS 695
OY 217 -----TKYKSSPEKSTASD-PYSDEKVTGRIDKNVSPARPPLVAIPIVHYDM 266

```

```

Db 696 AKESIDAKYTEITETKLETFNKKDKDYKEKELEKTRKIDDEFINTREINP----- 743
OY 267 ENILSKNDOSTONTDSEFTISKNTSTKSTHSEVGNMVAHASFIDIGGSVAGSNGN 326
OY 744 -----NSTLISELTSKSDSKNSVTSNSKSDI----- 771
OY 327 SNSSTVAIDHSISLAG-ERTMAETMGLNTADTARLANIRYNTCAPIYNNVLPITSLVL 385
Db 772 -ETANTELQALAKANDKQADNLAKSTKE--OLNNSISSANLILAKLID----- 819
OY 386 GKNOTLATIKA---KENQISQILAPNNYPSKNLAPALNAQ-DIFSSTPIYNNVQFLE 441
Db 820 -KDNITQAKTELEKEVOKANQAVASNNTASQSKSLAKYTEITKLETFNKKDKVK 878
OY 442 LEKTRQRLDTPDQVGNATATYNGENRVAVDTSNWSEVLPOIOTFARLITFNKIDILNV 501
Db 879 FRELQTRKIDDEF-----INTKTNPDYSTLISELTSK-----R 913
OY 502 ERRIAVNPSPDPLETTKPDMLKALKIAGFENGNGMLQYOKODITPEDFPDQOTSQ 561
Db 914 DSKNSITSSKSDIETANTLQALAKA-----NIDKQAD-NLARSTREQ 959
OY 562 IKKQIAELNATNIYTVLDRKIKLAKMLILRDKREHYDRNNIAVGA---ESVKEAARE 618
Db 960 LKNSISSAN---TLAKLT-----DKDNTIQAKTELEKEVOKANQA 998
OY 619 VINSSTEGILLNIDKDIRKILSGYIYEIPEDEGLKEVINDR-YDMLNITSLRQDGFID 677
Db 999 VASNNTASW---QSKSSLDKAVTEL--TKLETFNKKDKVATFRELQTRKIDDEFIN 1051
OY 678 FKYKNDKLPLVISNPNKYVNYAVT---KENTIDNPSNGTST? 718
Db 1052 TNK-----TNPNSTLISELTSKSDSKNSITNSKSDIER 1087

```

Search completed: October 4, 2002, 14:35:21
 Job time: 66 sec

Fri Oct 4 14:31:04 2002

us-09-848-909-21.rpt